AD-A249 623

I PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1991	THESIS/[illegible]

4. TITLE AND SUBTITLE
Maximum Likelihood Estimation of Fractional Brownian Motion and Markov Noise Parameters

5. FUNDING NUMBERS

6. AUTHOR(S)
Matthew E. Skeen, 2d Lt

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFIT Student Attending: Massachusettes Institute of Technology

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/CI/CIA- 91-132

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFIT/CI
Wright-Patterson AFB OH 45433-6583

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release IAW 190-1
Distributed Unlimited
ERNEST A. HAYGOOD, Captain, USAF
Executive Officer

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

14. SUBJECT TERMS

15. NUMBER OF PAGES
138

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18
298-102

132

Title: Maximum Likelihood Estimation of Fractional Brownian Motion and Markov Noise Parameters
Author: Matthew Edward Skeen
Military Rank: Second Lieutenant
Service Branch: Air Force
Date: 1991
Number of Pages: 138
Degree Awarded: Master of Science in Aeronautics and Astronautics
Institution: Massachusettes Institute of Technology

ABSTRACT

Maximum likelihood estimation and power spectral density analysis are developed as tools for the analysis of stochastic processes. Some useful results from the theory of Markov stochastic processes are then presented followed by the introduction of fractional Brownian motion and fractional Gaussian noise as non-Markov models for systems with power spectral density proportional to f^{β}, where $-3 < \beta < -1$ and $-1 < \beta < 1$ over all frequencies.

Maximum likelihood system identification is applied to estimating the unknown parameters in a Markov model which approximates fractional Brownian motion. The algorithm runs a Kalman filter on states and a maximum likelihood estimator on parameters. Results are presented from estimating trend, white noise, random walk, and exponentially correlated noise parameters from fits to simulated and real test data.

Maximum likelihood estimation is applied to the batch estimation of parameters in the non-Markov fractional Brownian motion model. New in this thesis is the use of partial derivatives to minimize the resulting likelihood function, and the capability to estimate the unknown parameters of additional trend and Markov noise processes. Results are presented from fits to computer simulated sample paths.

PRIMARY REFERENCES

Lundahl, Torbjörn, William J. Ohley, Steven M. Kay and Robert Siffert, "Fractional Brownian Motion: A Maximum Likelihood Estimator and Its Application to Image Texture," *IEEE Transactions on Medical Imaging*, Vol. MI-5, No. 3, September 1986, pp. 152-161.

Mandelbrot, Benoit B. and John W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, Vol. 10, 1969, pp. 422-437.

Sandel, Nils R. Jr. and Khaled I. Yared, "Maximum Likelihood Identification of State Space Models for Linear Dynamic Systems," M.I.T. Electronic Systems Laboratory Report ESL-R-814, Cambridge, MA, April 1978.

MAXIMUM LIKELIHOOD ESTIMATION OF FRACTIONAL BROWNIAN MOTION AND MARKOV NOISE PARAMETERS

by

MATTHEW E. SKEEN

B.S., Astronautical Engineering, United States Air Force Academy
Colorado Springs, Colorado (1990)

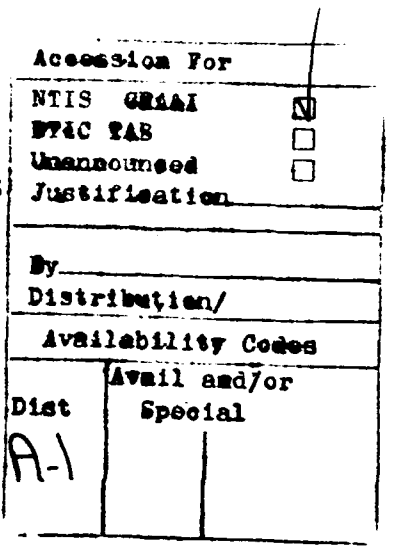
Submitted to the Department of Aeronautics and Astronautics
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE in AERONAUTICS AND ASTRONAUTICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1992



Signature of Author _____
Department of Aeronautics and Astronautics
20 December 1991

Approved by _____
Dr. Michael E. Ash
Principal Member Technical Staff, Charles Stark Draper Laboratory
Thesis Supervisor

Certified by _____
Professor Wallace E. Vander Velde
Department of Aeronautics and Astronautics
Thesis Advisor

Accepted by _____
Professor Harold Y. Wachman
Chairman, Department Graduate Committee

MAXIMUM LIKELIHOOD ESTIMATION OF FRACTIONAL BROWNIAN MOTION AND MARKOV NOISE PARAMETERS

by

MATTHEW EDWARD SKEEN

Submitted to the Department of Aeronautics and Astronautics
on December 20, 1991
in partial fulfillment of the requirements for the degree of
Master of Science in Aeronautics and Astronautics

ABSTRACT

Maximum likelihood estimation and power spectral density analysis are developed as tools for the analysis of stochastic processes. Some useful results from the theory of Markov stochastic processes are then presented followed by the introduction of fractional Brownian motion and fractional Gaussian noise as non-Markov models for systems with power spectral density proportional to f^{β}, where $-3 < \beta < -1$ and $-1 < \beta < 1$ over all frequencies.

Maximum likelihood system identification is applied to estimating the unknown parameters in a Markov model which approximates fractional Brownian motion. The algorithm runs a Kalman filter on states and a maximum likelihood estimator on parameters. Results are presented from estimating trend, white noise, random walk, and exponentially correlated noise parameters from fits to simulated and real test data.

Maximum likelihood estimation is applied to the batch estimation of parameters in the non-Markov fractional Brownian motion model. New in this thesis is the use of partial derivatives to minimize the resulting likelihood function, and the capability to estimate the unknown parameters of additional trend and Markov noise processes. Results are presented from fits to computer simulated sample paths.

Thesis Supervisor: Dr. Michael E. Ash
Principal Member Technical Staff, C. S. Draper Laboratory

Thesis Advisor: Professor Wallace E. Vander Velde
Department of Aeronautics and Astronautics

Acknowledgements

My sincerest gratitude is extended to my thesis supervisor, Dr. Michael E. Ash, not only for his enthusiastic assistance in my research, but also for his interest in my overall education. I would like to thank my thesis advisor, Prof. Wallace E. Vander Velde, for his guidance in the preparation of this document and for the care with which he reviewed the drafts.

I thank my friends at Draper Laboratory for all the help that they have given me and for making my time at MIT enjoyable. I give special thanks to Tanya for her love, support, and perseverance over the last 17 months.

This thesis was researched and written at the Charles Stark Draper Laboratory under Corporate Sponsored Research Project C68.

Publication of this thesis does not constitute approval by the laboratory of the findings or conclusions contained herein. It is published for the exchange and stimulation of ideas.

_____

Matthew E. Skeen

Table of Contents

List of Illustrations

List of Tables

Notation

Lower case letters denote scalars, underscored lower case letters denote vectors, and capital letters generally denote matrices.

A	=	state transition matrix
A_{ij}	=	element of the Hessian of negative log-likelihood function or Fisher Information approximation to the Hessian
B	=	system deterministic input matrix
B_{ij}	=	element of gradient of negative log-likelihood function
b	=	signal bias or random walk standard deviation
C	=	system output matrix
c_1	=	inverse of exponentially correlated noise time constant
c_2	=	exponentially correlated noise scaling parameter
$\det(\,)$	=	determinant operator
$d\beta$	=	Brownian motion differential
$E\{\,\}$	=	expected value operator
H	=	fractional Brownian motion dimension parameter
I	=	Fisher Information matrix
i	=	$\sqrt{-1}$ when not used as an index
K	=	Kalman filter gain matrix
L	=	system plant noise input matrix
$\ln[\,]$	=	natural logarithm operator
$\min(t_1,t_2)$	=	minimum value of t_1 and t_2
$N(\mu,\sigma^2)$	=	normally distributed with mean μ and variance σ^2
$Pr(\bullet)$	=	probability of the event $\bullet$
$p(\underline{x};\underline{\alpha})$	=	probability density of $\underline{x}$ as a function of the parameters $\underline{\alpha}$
$p(z_1,\ldots,z_N;\underline{\alpha})$	=	probability density function of the random variables $z_1, \ldots, z_N$ as a function of the parameters $\underline{\alpha}$

$p(\underline{x} \mid \underline{y};\underline{\alpha})$ = conditional probability density of $\underline{x}$ given the random variable $\underline{y}$ as a function of the parameters $\underline{\alpha}$

$p(\underline{z}(t_k) \mid \underline{z}^{k-1})$ = probability density of $\underline{z}(t_k)$ given $\underline{z}(t_{k-1})$, $\underline{z}(t_{k-2})$, ..., $\underline{z}(t_o)$

$\underline{r}$ = Kalman filter pre-update residual

S = covariance matrix of a multivariate normal distribution

$tr[\,]$ = trace operator

t_k = time index

$\underline{u}$ = deterministic input vector

V_H = variance of unscaled fractional Brownian motion

X = Fourier transform of x(t)

$\underline{x}$ = state vector or vector valued stochastic or deterministic process

$\hat{\underline{x}}$ = optimal estimate of state vector

$\underline{z}$ = measurement vector $[z_1,...,z_N]^T$

$\underline{\alpha}$ = vector of unknown system parameters

$\tilde{\underline{\alpha}}$ = parameter estimate

$\hat{\underline{\alpha}}$ = maximum likelihood estimate of parameters

β = Brownian motion process

β_H = fractional Brownian motion process

$\Gamma()$ = gamma function

Δt = discrete time increment

$\Delta \underline{y}(t_k)$ = increment between measurements at time t_k and at time t_{k-1}

$\Delta \underline{\alpha}$ = parameter adjustment

$\delta(t)$ = delta (impulse) function

$\delta(j)$ = discrete form of delta function

δ_{jk} = Kronecker delta (1 if j = k, 0 otherwise)

ζ = negative log-likelihood function (sometimes without constant)

Θ = covariance of discrete white measurement noise

$\underline{\theta}$ = white measurement noise vector

$\underline{\mu}$ = mean of measurements

Ξ = covariance of discrete white plant noise

$\underline{\xi}$ = white plant noise vector

Σ = covariance matrix

σ	=	standard deviation of a random variable
σ_H	=	standard deviation of fractional Brownian motion
Φ_{xx}	=	two-sided power spectral density of $\underline{x}$
Φ^1_{xx}	=	one-sided power spectral density of $\underline{x}$
$\tilde{\Phi}_{xx}$	=	estimate of power spectral density function
$\tilde{\Phi}^1_{xx}$	=	estimate of one-sided power spectral density function
ϕ_{xy}	=	correlation function of stationary processes $\underline{x}$ and $\underline{y}$
$\tilde{\phi}_{xy}$	=	estimate of correlation function
$\phi_{xy}(t_k, t_{k+j})$	=	correlation function of nonstationary processes $\underline{x}$ and $\underline{y}$
ω	=	probability space
*	=	convolution operator
$\lvert\ \rvert$	=	magnitude operator

Subscripts

i	=	i^{th} element of a vector
ij	=	ij^{th} element of a matrix

Superscripts

T	=	transpose operator
*	=	complex conjugate
-1	=	matrix inverse

Acronyms

fBm	=	fractional Brownian motion
FIMLOF	=	Full Information Maximum Likelihood Optimal Filtering
PSD	=	Power Spectral Density

Chapter 1

Introduction and Summary

1.1 Modeling Noise Processes and Estimating Noise Parameters

In this thesis, maximum likelihood estimation is applied to estimating stochastic noise parameters for both Markov and non-Markov noise processes. Dynamic parameters, such as trend, were simultaneously estimated with and separated from stochastic parameters. The goal was to develop methods of determining unknown parameters in models of systems which exhibit power spectral densities proportional to f^{β}, where $-2 \leq \beta \leq 0$.

There are many naturally occurring systems with PSD proportional to f^{β}. This behavior can persist over a wide frequency range [44]. The particular value $\beta = -1$ is indicative of what is called flicker or $\frac{1}{f}$ noise. Since no simple Markov noise process can have this frequency domain characteristic for more than a narrow frequency band, these systems are commonly approximated using a combination of several independent Markov processes [21].

An alternate approach is to model such a noise process using fractional Brownian motion [26], which has PSD proportional to f^{β} $(-3 < \beta < -1)$ for all frequencies, where β is determined by a parameter of the process. The increment process of fractional Brownian motion, called fractional Brownian noise, has PSD proportional to f^{β} $(-1 < \beta < 1)$ for all frequencies.

Because the Markov approximation to fractional Brownian motion can be implemented in state space form, this model is commonly used in applications where the goal is to develop a control law for a system [21]. Due

to its minimum parameter form for representing f^{β} noise, the fractional Brownian motion model is more frequently used in applications where the goal is to analyze a given signal, e.g. image processing [24].

Even though fractional Brownian motion (fBm) has stationary self-similar increments, they are not independent and fBm is not a Markov process. This means that state space models and Kalman filter estimators cannot be applied to the parameters of the process.

A Kalman filter estimator could be applied to estimating the stochastic and dynamic parameters of a Markov process modeled with a state dynamic system. The parameters would be estimated along with the states by augmenting the state vector and using an extended Kalman filter (the augmented system will be non-linear, because the dynamic parameters generally multiply the states, even if the original system is linear in the states). This is a somewhat artificial approach which might not converge [41].

As an alternate approach, maximum likelihood system identification can be applied to estimating the stochastic and dynamic parameters in a state dynamic system in conjunction with using a Kalman filter to estimate the states. The maximum likelihood technique can also be applied to estimating the parameters in a non-Markov noise process such as fractional Brownian motion simultaneously with other dynamic and Markov noise parameters. The classical Fisher maximum likelihood approach has many desirable properties, and is naturally applied to these problems. The results of this thesis show that it can be applied successfully.

New in this thesis is the use of partial derivatives in estimating fractional Brownian motion parameters, rather than a brute force search for a maximum of the likelihood function. Also new in this thesis is the estimation of other noise parameters and trend simultaneously with the fractional Brownian motion parameters. In addition, trend, white noise, random walk, and exponentially correlated noise parameters are estimated for a Markov state dynamic system model. This thesis provides a concrete example of this generally accepted procedure.

1.2 Background Material

1.2.1 Maximum Likelihood Estimation

Chapter 2 discusses the properties of parameter estimators in general and maximum likelihood estimators in particular. For any estimator, the Cramer-Rao lower bound applies with the covariance of unbiased parameter estimates being greater than or equal to the elements of the inverse of the Fisher information matrix [45]. The Fisher information matrix is the expected value of the Hessian of second partial derivatives of the negative log-likelihood function, where the likelihood is the probability density of the observables as functions of the parameters [45]. Rigorous calculations show that the Fisher information matrix is also equal to the expected value of the dyadic product of the gradient of the negative log-likelihood, which only involves first partial derivatives [45].

Maximum likelihood estimation seeks parameter values which maximize the likelihood function (or minimize the negative log likelihood), which means that parameter values are chosen that make it most likely that the observations that did occur would have occurred. Under many circumstances, maximum likelihood estimates have desirable theoretical properties, such as being asymptotically consistent, unbiased, efficient, normally distributed about the true parameter values, and attaining the Cramer-Rao lower bound [45].

Determination of maximum likelihood estimates in non-linear problems is done iteratively starting from a first guess for the parameter values. The adjustments to the parameter values at each stage of the iteration are determined by solving a set of linear equations whose coefficient matrix is the Fisher information approximation to the Hessian of the negative log-likelihood function.

1.2.2 Power Spectral Density Analysis

Chapter 3 discusses stochastic processes, the autocorrelation function of a stochastic process, and the Power Spectral Density (PSD) which is defined as the Fourier transform of the autocorrelation function.

The autocorrelation is the expected value of the product of the stochastic process values at two different times. For a stationary stochastic process, the autocorrelation is only a function of the difference of the two times. For an ergodic stationary stochastic process, expected value ensemble averages can be replaced by time averages over any realization of the process. Therefore, an estimate of the autocorrelation function of an ergodic stochastic process is the time domain autocorrelation over a finite interval of a particular realization from the process, divided by the time interval [34].

The Fourier transform of this estimate of the autocorrelation function is an estimate of the PSD of the process. By the frequency domain properties of time autocorrelation (Appendix A), the estimate of the PSD is then equal to the magnitude squared of the Fourier transform of a finite time span of the particular realization of the stochastic process. Finally, this estimate of the PSD can be computed using the discrete fast Fourier transform of samples from the realization [34].

PSD analysis is employed as an adjunct to the time domain estimation of stochastic noise parameters. It is checked that sample paths generated using random number generators and employing the estimated parameters have the same PSD as that of the original data.

1.2.3 Markov Noise Processes

Chapter 4 discusses the Markov and Martingale properties for stochastic processes, the Weiner random walk and white noise processes, stochastic integrals and differential equations, exponentially correlated noise, and ways of simulating noise processes.

The Markov property is that the expected value of the future given the present and the past is equal to the expected value of the future given the present [18]. This is an expression of the principle of causality from physics for stochastic processes. In-so-far as non-Markov behavior is seen in nature (such as f^{β} noise over a wide frequency band with $-2 < \beta < 0$), one could suppose that it is due to not having enough states in the model of the natural system. However it may be inconvenient to have as many states as would be required to practically model the noise.

A martingale is a stochastic process for which the expected value of the future is equal to the present [13]. A Weiner random walk process $\beta(t)$ is a Gaussian process with stationary independent increments. It is a martingale with $E\{[\beta(t_2)-\beta(t_1)]^2\} = \sigma^2 | t_2-t_1 |$, and it provides a model of the Brownian motion that a small particle buffeted by fluid molecules undergoes. It is almost surely continuous and non-differentiable [18].

White noise can be considered as the approximate derivative of a Weiner process. The 0 and -2 log-log PSD slopes of white noise and random walk noise are derived. The -2 and +2 log-log PSD slopes of trend and quantization noise are also derived.

The Ito stochastic integral is defined relative to the differential $d\beta$ of a Weiner process. The differential in terms of increments of the process can be used, even though the derivative does not exist. The independent increments properties of the Weiner process are heavily used, so that, for example, the stochastic integral could not be defined relative to the differential $d\beta_H$ of a fractional Brownian motion process. Results about the integration of stochastic differential equations are stated.

The stochastic differential equation for exponentially correlated noise is given, and its PSD is computed. The log-log PSD slope is 0 at low frequencies transitioning to -2 at high frequencies, so that a number of exponentially correlated noise processes could be used to approximate a -1 log-log PSD slope over a finite frequency band.

A technique is described for simulating these various noise processes using a random number generator on a computer.

1.2.4 Fractional Brownian Motion

Fractional Brownian motion $\beta_H(t)$ is defined in terms of a certain stochastic integral relative to a Weiner process differential $d\beta$. The integrand is in fact that used in the ordinary calculus definition of fractional derivatives and integrals. Thus fractional Brownian motion can be considered as the fractional derivative or integral of ordinary Brownian motion, depending on the value of the parameter H in the definition ($0 < H < 1$). It reduces to ordinary Brownian motion when $H = \frac{1}{2}$.

Fractional Brownian motion has stationary and self-similar but not independent increments with $E\{[\beta_H(t_2)-\beta_H(t_1)]^2\} = \sigma_H^2 \mid t_2-t_1 \mid^{2H}$. It is unique, in that any stochastic process with these properties is fractional Brownian motion multiplied by a constant.

The autocorrelation function of fractional Brownian motion (fBm) is calculated, from which it is derived that its log-log PSD slope is (-1-2H). This is done by showing that the fBm derivative process has log-log PSD slope (1-2H).

Two techniques for generating simulated fractional Brownian motion sample paths using a random number generator are presented.

1.3 Summary of Maximum Likelihood Fits to Real and Simulated Data

1.3.1 Maximum Likelihood System Identification for Markov Processes

Chapter 6 discusses maximum likelihood system identification for a state dynamic system model, and applies the technique to estimating trend, white noise, random walk, and exponentially correlated noise parameters in real and simulated data.

Maximum likelihood system identification runs a maximum likelihood estimator on dynamic and stochastic parameters and a Kalman

filter on the states. This technique is sometimes called Full Information Maximum Likelihood Optimal Filtering (FIMLOF).

The general system identification formulas are presented for a discrete linear dynamic system with a vector observable, and then for the specific case discussed in this thesis of a scalar observable that is the sum of a trend, white noise, random walk, and exponentially correlated noise. The system identification formulas were coded in computer software to run a Kalman filter on the states given a first guess to the parameters, adjust the values of the parameters using the partial derivatives of the log-likelihood function which are generated in running the Kalman filter, and repeat the process until convergence is obtained to the maximum likelihood estimates of the stochastic and dynamic parameters. A Cramer-Rao lower bound for the uncertainty of the parameter estimates is provided by the inverse of the Fisher information matrix which is calculated during this process.

The results were that the FIMLOF method provides an accurate estimate of Markov noise parameters if a sufficient number of measurements are available. The accuracy of such estimates was verified by comparing the Cramer-Rao lower bound to the estimation error in the case of computer generated sample paths where the "true" parameter values were available. When the method was applied to experimental data displaying a -1 log-log PSD slope, the frequency response characteristics of the model using Markov noise parameter estimates matched those of the real system.

1.3.2 Maximum Likelihood Estimation of Fractional Brownian and Other Parameters

In Chapter 7, the likelihood function for the sum of fractional Brownian and other Gaussian noise processes is computed from the autocorrelation functions of the processes, for both the increment and sum of increment formulations. The iterative determination of trend, fractional Brownian, and Markov noise parameters was coded on a computer using this expression for the likelihood function with an $N \times N$ measurement covariance matrix, where N is the number of observations. Fits to simulated

data were done with $N = 128$ and with $N = 200$. There are obvious computational difficulties for very large N. The non-Markov nature of fractional Brownian motion, with correlations extending over all observations, makes this batch approach rather than a sequential approach necessary.

Using this approach, it was possible to accurately estimate both stochastic and deterministic parameters in cases involving combinations of computer generated sample paths of fractional Brownian motion, trend, and white noise. The accuracies of these estimates were verified by comparing the Cramer-Rao lower bound to the estimation errors. However, when fractional Brownian motion was combined with exponentially correlated noise, $N = 200$ measurements were insufficient to allow the algorithm to converge. At this point, the need for more measurements came into conflict with the computational expense of inverting the $N \times N$ measurement covariance matrix and calculating its partial derivatives.

A successful estimation of fractional Brownian motion parameters was also made using the -1 log-log PSD slope experimental data that had been used in FIMLOF estimation.

Chapter 2

Maximum Likelihood Estimation

2.1 Parameter Estimation

The goal of this thesis is to estimate the values of unknown parameters in the models of fractional Brownian motion and Markov noise processes. The information that is available for use in this estimate is a set of measurements of the output of the model. There are several methods which are available for use in problems of this sort. Maximum likelihood estimation is chosen because of its simplicity and its well documented favorable properties which will be presented in this chapter.

2.2 Likelihood Function and Negative Log-Likelihood Function

Let $\underline{z} = [z_1, ..., z_N]^T$ be observations or measurements at times $t_1, ..., t_N$ of some system involving dynamic and noise processes dependent on parameters $\underline{\alpha} = [\alpha_1, ... , \alpha_q]^T$. The likelihood function is the joint probability density of the measurements as a function of the measurement and parameter values:

$$p(\underline{z};\underline{\alpha}) = p(z_1, ... , z_N ; \alpha_1, ... ,\alpha_q) \qquad (2.2\text{-}1)$$

Since $p(\underline{z};\underline{\alpha})$ is a probability density,

$$\int p(\underline{z};\underline{\alpha})\, d\underline{z} \;=\; 1 \qquad (2.2\text{-}2)$$

Taking partial derivatives of both sides of the above equation yields

$$\int \frac{\partial p(\underline{z};\underline{\alpha})}{\partial \alpha_i}\, d\underline{z} \;=\; 0, \quad \int \frac{\partial^2 p(\underline{z};\underline{\alpha})}{\partial \alpha_i \partial \alpha_j}\, d\underline{z} \;=\; 0 \qquad (2.2\text{-}3)$$

The negative log-likelihood is the negative of the natural logarithm of the likelihood function:

$$\zeta(\underline{z};\underline{\alpha}) \;=\; -\ln[p(\underline{z};\underline{\alpha})] \qquad (2.2\text{-}4)$$

The maximum likelihood estimates for the parameters $\underline{\alpha}$ are the values that maximize the likelihood function or equivalently minimize the negative log-likelihood function given a set of measurements $\underline{z}$.

2.3 Fisher Information Matrix

The Fisher information matrix is the expected value of the Hessian of second partial derivatives of the negative log-likelihood function:

$$I_{ij} \equiv E\left\{ \frac{\partial^2 \zeta(\underline{z};\underline{\alpha})}{\partial \alpha_i \partial \alpha_j} \right\} \qquad (2.3\text{-}1)$$

where E{ } denotes expectation (integration over the probability density of the random variable). As will be explained in the sequel, this matrix provides a measure of the information contained in a parameter estimate.

It will be useful to have a simplified expression for the information matrix which does not involve second partial derivatives. By Equations (2.2-3) and (2.2-4)

$$I_{ij} \;=\; -\int \frac{\partial^2 \ln[p(\underline{z};\underline{\alpha})]}{\partial \alpha_i \partial \alpha_j}\, p(\underline{z};\underline{\alpha})\, d\underline{z}$$

$$\;=\; -\int \frac{\partial}{\partial \alpha_i}\left[\frac{1}{p(\underline{z};\underline{\alpha})} \frac{\partial p(\underline{z};\underline{\alpha})}{\partial \alpha_j}\right] p(\underline{z};\underline{\alpha})\, d\underline{z}$$

$$= -\int \frac{\partial^2 p(\underline{z};\underline{\alpha})}{\partial\alpha_i \partial\alpha_j}\, d\underline{z}$$

$$+\int \frac{1}{p(\underline{z};\underline{\alpha})^2} \frac{\partial p(\underline{z};\underline{\alpha})}{\partial\alpha_i} \frac{\partial p(\underline{z};\underline{\alpha})}{\partial\alpha_j}\, p(\underline{z};\underline{\alpha})\, d\underline{z}$$

$$= E\left\{ \frac{\partial \ln[p(\underline{z};\underline{\alpha})]}{\partial\alpha_i} \frac{\partial \ln[p(\underline{z};\underline{\alpha})]}{\partial\alpha_j} \right\}$$

$$= E\left\{ \frac{\partial\zeta(\underline{z};\underline{\alpha})}{\partial\alpha_i} \frac{\partial\zeta(\underline{z};\underline{\alpha})]}{\partial\alpha_j} \right\} \qquad (2.3\text{-}2)$$

2.4 Cramer-Rao Lower Bound

An estimate $\tilde{\alpha}_j$ of a parameter α_j is a function of the measurements:

$$\tilde{\alpha}_j = f_j(z_1, ..., z_N) \qquad (2.4\text{-}1)$$

As a function of the measurement random variables it is also a random variable. The function f_j cannot be arbitrarily chosen if the parameter estimate is to have good properties. One desirable property of an estimator is that it be unbiased:

$$E\{\tilde{\alpha}_j\} = \text{true value of } \alpha_j \qquad (2.4\text{-}2)$$

Another desirable property is that its variance be small, although there is a lower bound as to how small it can be.

In the case of a scalar parameter, if $\tilde{\alpha}$ is an estimate of the parameter α then [7]

$$\int (\tilde{\alpha} - \alpha)\, p(\underline{z};\alpha)\, d\underline{z} = b(\alpha) \qquad (2.4\text{-}3)$$

where $b(\alpha)$ is the bias associated with the estimate. Assuming that $p(\underline{z};\alpha)$ has a first derivative and taking the partial derivative of each side gives

$$\int \frac{\partial}{\partial \alpha} [(\tilde{\alpha} - \alpha)\, p(\underline{z};\alpha)]\, d\underline{z} = b'(\alpha) \qquad (2.4\text{-}4)$$

where $b'(\alpha)$ is the derivative of the bias with respect to α. Taking the derivative of the product gives

$$- \int p(\underline{z};\alpha)\, d\underline{z} + \int (\tilde{\alpha} - \alpha) \frac{\partial}{\partial \alpha} p(\underline{z};\alpha)\, d\underline{z} = b'(\alpha) \qquad (2.4\text{-}5)$$

$$\int (\tilde{\alpha} - \alpha) \{ \frac{\partial}{\partial \alpha} \ln[p(\underline{z};\alpha)] \}\, p(\underline{z};\alpha)\, d\underline{z} = 1 + b'(\alpha) \qquad (2.4\text{-}6)$$

At this point, the Schwarz inequality for integration with respect to the measure $p(\underline{z};\alpha)\, d\underline{z}$ may be applied to give

$$\int (\tilde{\alpha} - \alpha)^2\, p(\underline{z};\alpha)\, d\underline{z} \int \{ \frac{\partial}{\partial \alpha} \ln[p(\underline{z};\alpha)] \}^2\, p(\underline{z};\alpha)\, d\underline{z} \geq [1 + b'(\alpha)]^2 \qquad (2.4\text{-}7)$$

Finally, using the definition of the expected value operator and the scalar form of Equation (2.3-2) leaves a lower bound for the variance of an estimate:

$$E\{(\tilde{\alpha} - \alpha)^2\} \geq \frac{[1 + b'(\alpha)]^2}{I} \qquad (2.4\text{-}8)$$

This lower bound is known as the Cramer-Rao lower bound. Equation (2.4-8) shows that for an unbiased scalar estimator, the Cramer-Rao lower bound is equal to the inverse of the Fisher information matrix, which has specialized to a scalar.

In the case of an unbiased estimate of q unknown parameters, this lower bound generalizes to [45]

$$E\{(\tilde{\alpha} - \alpha)^2\} - I^{-1} \geq 0 \qquad (2.4\text{-}9)$$

which means that every element of the covariance of the estimate must be greater than the corresponding element of the inverse of the Fisher information matrix. Thus the Cramer-Rao lower bound for the variance of the i^{th} parameter estimate is equal to $[I^{-1}]_{ii}$. In this manner, the Fisher information matrix provides a measure of the amount of information an

estimator employs by providing a limit to the accuracy with which each parameter may be estimated.

2.5 Properties of Maximum Likelihood Estimates

As stated in Section 2.2, maximum likelihood estimates $\hat{\underline{\alpha}} = [\hat{\alpha}_1, ..., \hat{\alpha}_q]^T$ of the parameters are those for which the likelihood function $p(\underline{z};\hat{\underline{\alpha}})$ is a maximum, or equivalently for which the negative log-likelihood $\zeta(\underline{z};\hat{\underline{\alpha}})$ is a minimum. In other words, maximum likelihood estimated parameter values are such that it is most likely that the observations that did occur would have occurred.

In order for maximum likelihood estimation to be carried out, it is necessary that for two different parameter vectors $\underline{\alpha}'$ and $\underline{\alpha}''$, the joint probability density of the observables that occurred should not be the same:

$$p(\underline{z};\underline{\alpha}') \neq p(\underline{z};\underline{\alpha}'') \qquad (2.5\text{-}1)$$

As stated at the start of this chapter, one reason for using maximum likelihood estimation in this thesis was the fact that it has favorable properties. These properties are that as the number of measurements approaches infinity, the maximum likelihood estimate is consistent, efficient, and sufficient. These properties are explained in the following paragraphs.

If an estimate $\tilde{\underline{\alpha}}$ converges in probability to the true value $\bar{\underline{\alpha}}$ of a parameter vector $\underline{\alpha}$ as the number of measurements $N \rightarrow \infty$, it is called a consistent estimate of $\underline{\alpha}$ [45]. This means that the estimate is unbiased and the covariance of the estimate goes to zero as the number of measurements approaches infinity.

If an estimate $\tilde{\underline{\alpha}}$ is an unbiased estimate for $\bar{\underline{\alpha}}$ such that no other unbiased estimate has a smaller variance, then it is called an efficient estimate of $\bar{\underline{\alpha}}$ [45]. An asymptotically efficient estimate is efficient as the number of measurements becomes large.

If $\tilde{\underline{\alpha}}$ is an unbiased estimator for $\underline{\alpha}$ such that for any statistic $\tilde{\tilde{\underline{\alpha}}}$ (function of the observables $\underline{z}$), the distribution of the conditional random variable $\tilde{\tilde{\underline{\alpha}}} | \tilde{\underline{\alpha}}$ does not depend on $\underline{\alpha}$, then $\tilde{\underline{\alpha}}$ is a sufficient estimator for $\underline{\alpha}$[45].

If $\underline{z} = [z_1, ..., z_N]^T$ are independent measurement random variables with the same probability distributions $q(z_j;\underline{\alpha})$, so that the joint density is $p(\underline{z};\underline{\alpha}) = q(z_1;\underline{\alpha})\cdots q(z_N;\underline{\alpha})$, then the maximum likelihood estimates $[\hat{\alpha}_1, ..., \hat{\alpha}_q]$ are asymptotically consistent, efficient, sufficient, normally distributed about the true value $\underline{\alpha}$, and the Cramer-Rao lower bound becomes tight [45].

If the measurements are not independent identically distributed random variables, it can be shown [28] that the maximum likelihood estimates are still asymptotically consistent, unbiased, efficient, and normally distributed about the true value $\underline{\alpha}$. However, this requires that the first, second, and third partial derivatives of the likelihood function exist over the admissible range of parameters. Thus, maximum likelihood estimation is still attractive, and the Cramer-Rao lower bound remains a useful measure of the accuracy of the estimator in the limit of a large number of observations.

2.6 Iterative Determination of Maximum Likelihood Estimates

Let $\hat{\underline{\alpha}} = [\hat{\alpha}_1, ... , \hat{\alpha}_q]^T$ be the maximum likelihood estimates of the parameters $\underline{\alpha}$. For these values the likelihood function is a maximum, or the negative log-likelihood is a minimum:

$$\zeta(\underline{z};\hat{\underline{\alpha}}) = \text{minimum} \qquad (2.6\text{-}1)$$

where $\underline{z} = [z_1, ..., z_N]^T$ are the measurements. This means that

$$\left.\frac{\partial \zeta(\underline{z};\underline{\alpha})}{\partial \alpha_i}\right|_{\underline{\alpha}=\hat{\underline{\alpha}}} = 0 \quad , \; i = 1, ..., q \qquad (2.6\text{-}2)$$

Let $\underline{\alpha}_o = [\alpha_{1o}, ..., \alpha_{qo}]^T$ be first guesses for the values of the parameters, with

$$\Delta\alpha_j = \hat{\alpha}_j - \alpha_{jo} \quad , \; j = 1, .. , q \tag{2.6-3}$$

Given $\underline{\alpha}_o$ and the measurements $\underline{z} = [z_1, ..., z_N]^T$, the negative log-likelihood function $\zeta(\underline{z};\underline{\alpha}_o)$ is calculated along with its first and second partial derivatives with respect to the α_i in order to determine the adjustments $\Delta\alpha_j$ to the first guesses α_{jo}.

Approximating Equation (2.6-2) with the first two terms of a Taylor series expansion gives

$$\begin{aligned} 0 &= \frac{\partial\zeta(\underline{z};\underline{\alpha})}{\partial\alpha_i}\Bigg|_{\underline{\alpha}=\hat{\underline{\alpha}}} \\ &= \frac{\partial\zeta(\underline{z};\underline{\alpha})}{\partial\alpha_i}\Bigg|_{\underline{\alpha}=\underline{\alpha}_o} + \sum_{j=1}^{q} \frac{\partial^2\zeta(\underline{z};\underline{\alpha})}{\partial\alpha_i\partial\alpha_j}\Bigg|_{\underline{\alpha}=\underline{\alpha}_o} \Delta\alpha_j \end{aligned} \tag{2.6-4}$$

This leads to the following set of linear equations to solve for the update $\Delta\alpha_j$ to the guesses α_{jo} to approach the maximum likelihood estimates $\hat{\alpha}_j$:

$$\sum_{j=1}^{q} A_{ij}\,\Delta\alpha_j = B_i \quad , \; i = 1, ..., q \tag{2.6-5}$$

where

$$B_i = -\frac{\partial\zeta(\underline{z};\underline{\alpha})}{\partial\alpha_i}\Bigg|_{\underline{\alpha}=\underline{\alpha}_o} \quad , \; i = 1, ..., q \tag{2.6-6}$$

$$A_{ij} = \frac{\partial^2\zeta(\underline{z};\underline{\alpha})}{\partial\alpha_i\partial\alpha_j}\Bigg|_{\underline{\alpha}=\underline{\alpha}_o} \quad , \; i,j = 1, ..., q \tag{2.6-7}$$

The expected value of the coefficient matrix A_{ij} is the Fisher information matrix. As an approximation, A_{ij} can be replaced by its expected value, obtaining by Equations (2.3-1) and (2.3-2)

$$A_{ij} \cong E\left\{ \frac{\partial\zeta(\underline{z};\underline{\alpha})}{\partial\alpha_i} \frac{\partial\zeta(\underline{z};\underline{\alpha})]}{\partial\alpha_j} \right\} \quad , \; i,j = 1, ..., q \tag{2.6-8}$$

With the Fisher information approximation to the Hessian of the negative log-likelihood function, only the gradient vector of first partial derivatives of the negative log-likelihood function need be calculated. In

some cases in which both the exact and approximate Hessians were calculated in this thesis, the Fisher information approximation gave better iterative convergence to the maximum likelihood parameter estimates than using the exact Hessian (see Section 6.6).

Using the new values of the parameters obtained by solving Equations (2.6-5), the negative log-likelihood and its partial derivatives are re-calculated and further adjustments $\Delta\alpha_j$ are made to the parameters. The iteration continues until convergence is obtained to the maximum likelihood estimates of the parameters.

2.7 Relation to Least Squares Estimates

Suppose the measurements z_k are related to theoretical model functions f_k by

$$z_k = f_k(t_k;\underline{\alpha}) + \varepsilon_k \quad , \quad k=1,...,N \tag{2.7-1}$$

where ε_k are independently normally distributed zero mean random errors with standard deviations δ_k. The likelihood function is

$$p(\underline{z};\underline{\alpha}) = \frac{1}{(2\pi)^{N/2}\,\delta_1\cdots\delta_N}\, e^{-\Sigma\,(z_k-f_k(t_k;\underline{\alpha}))^2/(2\delta_k^2)} \tag{2.7-2}$$

Maximum likelihood estimates of $\underline{\alpha}$ which maximize Equation (2.7-2) or minimize the negative log-likelihood

$$\zeta(\underline{z};\underline{\alpha}) = \left[\frac{N}{2}\ln(2\pi) + \ln(\delta_1\cdots\delta_N)\right] + \frac{1}{2}\sum_{k=1}^{N}\frac{(z_k-f_k(t_k;\underline{\alpha}))^2}{\delta_k^2} \tag{2.7-3}$$

are the same as weighted least squares estimates which minimize the Σ term in Equation (2.7-3), presuming that the constant part in brackets [ ] does not depend on $\alpha_1, ..., \alpha_q$.

The condition for a minimum is

$$\frac{\partial \zeta(\underline{z};\underline{\alpha})}{\partial \alpha_i} = \sum_{k=1}^{N} \frac{z_k - f_k(t_k;\underline{\alpha})}{\delta_k^2} \frac{\partial f_k(t_k;\underline{\alpha})}{\partial \alpha_i} = 0 \quad \text{when } \underline{\alpha} = \underline{\hat{\alpha}} \qquad (2.7\text{-}4)$$

The Fisher information matrix Equation (2.6-8) is

$$A_{ij} = \sum_{k=1}^{N} \frac{1}{\delta_k^2} \frac{\partial f_k(\underline{z};\underline{\alpha})}{\partial \alpha_i} \frac{\partial f_k(\underline{z};\underline{\alpha})}{\partial \alpha_j} \quad , \quad i,j = 1, \ldots, q \qquad (2.7\text{-}5)$$

since the partial derivatives of the f_k are non-random functions and the $z_k - f_k$ are independently normally distributed zero mean random variables with standard deviations δ_k.

The same set of linear equations for the adjustments to first guesses for the parameters as Equations (2.6-5), called normal equations in least squares estimation, can be obtained from Equation (2.7-4) with a first order Taylor expansion for f_k without any statistical assumptions or second partial derivatives. Namely, let $\underline{\alpha}_o = [\alpha_{1o}, \ldots, \alpha_{mo}]^T$ be first guesses for the values of the parameters, with Equation (2.6-3) giving the adjustments $\Delta\alpha_j$ towards the least squares estimates $\hat{\alpha}_j$. Assume that

$$f_k(\underline{z};\underline{\alpha}) = f_k(\underline{z};\underline{\alpha}_o) + \sum_{j=1}^{N} \frac{\partial f_k(\underline{z};\underline{\alpha}_o)}{\partial \alpha_j} \Delta\alpha_j \qquad (2.7\text{-}6)$$

$$\frac{\partial f_k(\underline{z};\underline{\hat{\alpha}})}{\partial \alpha_j} = \frac{\partial f_k(\underline{z};\underline{\alpha}_o)}{\partial \alpha_j} \qquad (2.7\text{-}7)$$

Inserting Equations (2.7-6) and (2.7-7) into (2.7-4) yields Equations (2.6-5), with A_{ij} being given by Equation (2.7-5).

Thus, least squares estimation involves the same calculations as maximum likelihood estimation, but if the statistics of the measurement errors ε_k are known, maximum likelihood theory allows the Cramer-Rao lower bound to be applied, with a lower bound for the covariance of the least squares, maximum likelihood estimates being the inverse of the coefficient matrix of the normal equations.

2.8 Relation to Other Estimators

The maximum likelihood method described in the previous sections is known as classical or Fisher maximum likelihood estimation. The state estimator used in a Kalman filter is called a Bayes estimator, because it involves conditional expectation. Namely, the Kalman filter estimate of the state $\underline{x}$ at a given time t_N in a state dynamic system is the expected value of the state given the observations $z_1, \ldots, z_N$ up to that time and the initial probability density of the states at time t_0.

If a state dynamic system involves unknown parameters $\underline{\alpha}$ which are to be estimated, then the state vector $\underline{x}$ could be augmented by states $\underline{\alpha}$ with dynamic equation $d\underline{\alpha} = 0 + \text{noise}$. However, the expanded state dynamic system is generally nonlinear (involving products of $\underline{x}$ and $\underline{\alpha}$), even if the original system equations were linear in $\underline{x}$. Nonlinear systems require the implementation of an extended Kalman filter. Convergence can be a problem for an extended Kalman filter estimate of $\underline{x}$ and $\underline{\alpha}$ [23].

The approach taken in Chapter 6 to estimate the parameters $\underline{\alpha}$ in a state dynamic system is to apply maximum likelihood estimation to the probability density $p(\underline{z};\underline{\alpha})$ of the observables $\underline{z}$ as a function of the parameters $\underline{\alpha}$ generated by running the Kalman filter to estimate the states assuming values for $\underline{\alpha}$. This probability density is not written $p(\underline{z} \mid \underline{\alpha})$, because no conditional probabilities are involved, as there is no prior probability information about $\underline{\alpha}$.

In Chapter 7 maximum likelihood estimation of $\underline{\alpha}$ is applied to a probability density $p(\underline{z};\underline{\alpha})$ arising from a fractional Brownian noise process, with no Kalman filter involved. It is a non-Markov process and the Kalman filter derivation depends on the Markov property (see Section 4.7).

The Bayesian maximum a posteriori probability (MAP) estimator is slightly different from the Fisher maximum likelihood estimator. Instead of maximizing the probability of the measurements as a function of the parameters, this method consists of maximizing the probability of the parameters as a function of the measurements [17]. This probability is found using Bayes' rule

$$p(\underline{\alpha} \mid \underline{z}) = \frac{p(\underline{z} \mid \underline{\alpha})\, p(\underline{\alpha})}{p(\underline{z})} \tag{2.8-1}$$

Note that the negative log-likelihood function of Equation (2.8-1) is

$$\zeta'(\underline{z};\underline{\alpha}) = -\ln[p(\underline{z} \mid \underline{\alpha})] - \ln[p(\underline{\alpha})] + \ln[p(\underline{z})] \tag{2.8-2}$$

The difference between this equation and Equation (2.2-4) is that Equation (2.8-2) requires the probability densities of $\underline{\alpha}$ and $\underline{z}$. If there is no prior knowledge of the distribution of the parameters, this distribution must be assumed uniform because all parameter values are equally likely. In addition, $p(\underline{z})$ must also be assumed uniform because this distribution is a function of the unknown parameters $\underline{\alpha}$. In this case, neither $p(\underline{\alpha})$ nor $p(\underline{z})$ is a function of the parameters themselves, and the MAP or Bayesian maximum likelihood estimate reduces to the classical Fisher maximum likelihood estimate.

Chapter 3

Power Spectral Density Analysis

3.1 Stochastic Processes

A stochastic, or random, process consists of a family of random variables which are taken from a probability space ω and indexed by a parameter t. These processes, which may be vector valued, are commonly written as $\underline{x}(t,\omega)$ or simply $\underline{x}(t)$. A random process is distinguished from a deterministic process by the fact that the latter is known exactly over the time span of interest while the former involves some element of chance.

A stochastic process may be placed into one of four categories based upon the associated probability space and index [18] The probability space may be either continuous or discrete, and the index may also be either continuous or discrete. Random variables drawn from a discrete probability space may only take discrete values though there may be an infinite number of possible values to choose from. The set of integers is an example of a discrete probability space. A continuous probability space allows random variables to take any value within a specified range. The set of real numbers is an example of a continuous probability space. This thesis will only consider signals defined on continuous probability spaces, i.e. systems with unknown parameters that may take any value within a given range. The distinction between continuous and discrete indices is analogous to that between continuous and discrete time systems. In fact, the index will be considered to be time throughout the remainder of this thesis.

A sample path or realization of a stochastic process is the time function $\underline{x}(t,\omega)$ for a fixed value of ω. Different sample paths of the same stochastic process will not, in general, be identical. The set of all possible sample paths of the process is called its ensemble.

3.2 Autocorrelation Function of a Stochastic Process

One way to characterize stochastic processes is to use correlation functions. The correlation function of two processes $\underline{x}(t)$ and $\underline{y}(t)$ is defined to be the expected value of their product at different times t_1, t_2 [8]:

$$\phi_{xy}(t_1,t_2) \equiv E\{\underline{x}(t_1)\,\underline{y}(t_2)^T\} \tag{3.2-1}$$

In general, ϕ_{xy} is a matrix with the element $[\phi_{xy}]_{ij}$ given by the scalar correlation function

$$[\phi_{xy}(t_1,t_2)]_{ij} = E\{x_i(t_1)\,y_j(t_2)\} \tag{3.2-2}$$

The units of a correlation function are equal to the product of the units of the signals of interest.

The correlation function of a process with itself is called its autocorrelation function and is given by

$$\phi_{xx}(t_1,t_2) = E\{\underline{x}(t_1)\,\underline{x}(t_2)^T\} \tag{3.2-3}$$

In this special case, the diagonal terms are actually the autocorrelations of the scalar processes which make up the vector $\underline{x}(t)$ and the off diagonal terms are known as the cross correlations between the individual scalar processes. The autocorrelation function of a process is essentially a measure of the dependence of the value of the process at one time with its value at other times.

The covariance function of a random process is found by subtracting the mean from the process and calculating the autocorrelation of the resulting signal:

$$\mathrm{cov}[\underline{x}(t_1,t_2)] = E\{\,[\underline{x}(t_1) - \underline{\mu}(t_1)][[\underline{x}(t_2) - \underline{\mu}(t_2)]^T\,\} \tag{3.2-4}$$

where

$$\underline{\mu}(t) = E\{\underline{x}(t)\} \tag{3.2-5}$$

The fact that the autocorrelation function of a zero mean stochastic process is equal to its covariance will be of use in Chapter 7.

A stochastic process is called stationary if its probability density is not a function of time [8]. This means that none of the statistics of the process will be functions of absolute time. A stationary random process is analogous to a time invariant deterministic system. The autocorrelation function of such a process will depend only upon the difference between the time indices $\tau = t_1 - t_2$. The autocorrelation function of a stationary process is even, i.e. $\phi_{xx}(\tau) = \phi_{xx}(-\tau)$. This may be illustrated using the change of variable $s = t + \tau$:

$$\phi_{xx}(\tau) = E\{x(t+\tau)x(t)\} = E\{x(s)x(s-\tau)\} = \phi_{xx}(-\tau) \tag{3.2-6}$$

Note also for a stationary process with $s = t + \frac{\tau}{2}$

$$\phi_{xx}(\tau) = E\{x(t+\tau)x(t)\} = E\{x(s + \tfrac{\tau}{2})x(s - \tfrac{\tau}{2})\} \tag{3.2-7}$$

An ergodic process is a stationary process that has the additional property that time averaging over a particular realization is equivalent to ensemble averaging, so that in particular

$$\phi_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t+\tau)\, x(t)\, dt \tag{3.2-8}$$

for any realization x(t) [6]. This is a useful property to have when dealing with experimental data because it is not common to have a large set of sample paths available for use in ensemble averaging.

3.3 Estimation of the Autocorrelation Function

For a scalar, ergodic, stochastic process, given a span of data x(t) where $-T \le t \le T$, a change of variables in Equation (3.2-8) shows that an estimator for the autocorrelation function is given by [34]

$$\tilde{\phi}_{xx}(\tau) = \frac{1}{2T} \int_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} x(t+\tfrac{\tau}{2})\, x(t-\tfrac{\tau}{2})\, dt \ , \ |\tau| < 2T \qquad (3.3\text{-}1)$$

It will be useful in Section 3.5 to note that the autocorrelation given in Equation (3.3-1) is similar to a convolution of x(t). Appendix A reviews the definitions of correlation and convolution and provides some useful relations.

If $p(\omega)$ is the probability density of the stationary, ergodic process x(t) then the expected value of the estimate (3.3-1) is given by [34]

$$E\{\tilde{\phi}_{xx}(\tau)\} = \int_{-\infty}^{\infty} \left[\frac{1}{2T} \int_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} x(t+\tfrac{\tau}{2})\, x(t-\tfrac{\tau}{2})\, dt \right] p(\omega)\, d\omega$$

$$= \frac{1}{2T} \int_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} \left[\int_{-\infty}^{\infty} x(t+\tfrac{\tau}{2})\, x(t-\tfrac{\tau}{2})\, p(\omega)\, d\omega \right] dt$$

$$= \frac{1}{2T} \int_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} \phi_{xx}(\tau)\, dt \qquad \text{by Equation (3.2-7)}$$

$$= \left(1 - \frac{|\tau|}{2T}\right) \phi_{xx}(\tau) \qquad (3.3\text{-}2)$$

Equation (3.3-2) shows that the estimate of the autocorrelation function which is given by Equation (3.3-1) is unbiased in the limit as T becomes very large with respect to τ [34]. Multiplying $\tilde{\phi}_{xx}(\tau)$ by $(1 - |\tau|/2T)^{-1}$ will also make the estimate unbiased.

3.4 Power Spectral Density of a Stationary Stochastic Process

The power spectral density (PSD) of a stationary random process is defined to be the Fourier transform of its autocorrelation function [8]

$$\Phi_{xx}(f) \equiv \int_{-\infty}^{\infty} \phi_{xx}(\tau)\, e^{-2\pi i f \tau}\, d\tau\,, \quad -\infty \le f \le \infty \tag{3.4-1}$$

where f is frequency in Hz and $i = \sqrt{-1}$. The PSD has units equal to (units^2)/Hz if the process x(t) is measured in units.

Because the autocorrelation function of a stationary stochastic process is even, the PSD is also even so that the one-sided PSD is equal to [5]

$$\Phi^{1}_{xx}(f) \equiv 2 \int_{-\infty}^{\infty} \phi_{xx}(\tau)\, e^{-2\pi i f \tau}\, d\tau\,, \quad f \ge 0 \tag{3.4-2}$$

$$\Phi^{1}_{xx}(f) = \Phi_{xx}(f) + \Phi_{xx}(-f)\,, \quad f \ge 0 \tag{3.4-3}$$

The one-sided PSD is often used in analyzing data where it is convenient to regard frequency f as being positive, whereas the two-sided PSD is more convenient for mathematical proofs.

As will be shown in Chapters 4 and 5, different types of stochastic processes have their own characteristic PSD shapes. This makes the PSD useful for determining the frequency range where a given process is more prominent in a given signal.

3.5 Estimation of the Power Spectral Density

3.5.1 Continuous Data

It is often desirable to determine the PSD of a stationary, ergodic, scalar stochastic process from a finite sample path of that process. One way to do this is to take the Fourier transform of the estimate of the autocorrelation function defined by Equation (3.3-1) [34]:

$$\tilde{\Phi}_{xx}(f) = \int_{-2T}^{2T} \tilde{\phi}_{xx}(\tau)\, e^{-2\pi i f \tau}\, d\tau \qquad (3.5.1\text{-}1)$$

or

$$\tilde{\Phi}_{xx}(f) = \int_{-2T}^{2T} \left[\frac{1}{2T} \int_{-T+\frac{|\tau|}{2}}^{T-\frac{|\tau|}{2}} x(t+\tfrac{\tau}{2})\, x(t-\tfrac{\tau}{2})\, dt \right] e^{-2\pi i f \tau}\, d\tau \qquad (3.5.1\text{-}2)$$

Note that this integral is only evaluated between -2T and 2T because the estimate of the autocorrelation function is only defined on this region and is assumed to be zero outside of it.

Taking advantage of the fact that $\tilde{\phi}_{xx}$ is an autocorrelation of x(t), assumed zero for $|t| > T$, Equation (A-6) in Appendix A shows that Equation (3.5.1-2) may be expressed as

$$\begin{aligned} \tilde{\Phi}_{xx}(f) &= \tfrac{1}{2T} X(f)\, X^*(f) \\ &= \tfrac{1}{2T} |X(f)|^2 \end{aligned} \qquad (3.5.1\text{-}3)$$

where X(f) indicates the finite Fourier transform of x(t), and X*(f) indicates its complex conjugate.

The finite Fourier transform of x(t) is given by

$$X(f) = \int_{-T}^{T} x(\tau)\, e^{-2\pi i f \tau}\, d\tau \qquad (3.5.1\text{-}4)$$

when x(t) is defined over the interval $-T \le t \le T$. This is equivalent to taking the integral from negative to positive infinity and setting x(t) equal to zero outside of the range defined by T.

The one sided PSD estimate is given by

$$\tilde{\Phi}^1_{xx}(f) = \tfrac{1}{T} |X(f)|^2 \ , \quad f \ge 0 \qquad (3.5.1\text{-}5)$$

The expected value of the estimate of the PSD is given by

$$E\{\tilde{\Phi}_{xx}(f)\} = \int_{-\infty}^{\infty} \left[\int_{-2T}^{2T} \tilde{\phi}_{xx}(\tau)\, e^{-2\pi i f \tau}\, d\tau \right] p(\omega)\, d\omega$$

$$= \int_{-2T}^{2T} E\{\tilde{\phi}_{xx}(\tau)\}\, e^{-2\pi i f \tau}\, d\tau$$

$$= \int_{-2T}^{2T} \left(1 - \frac{|\tau|}{2T}\right) \phi_{xx}(\tau)\, e^{-2\pi i f \tau}\, d\tau \qquad (3.5.1\text{-}6)$$

Equation (3.5.1-6) shows that the estimate of the PSD will be unbiased as T becomes very large, i.e., as the span of available data becomes very large:

$$\lim_{T\to\infty} E\{\tilde{\Phi}_{xx}(f)\} = \int_{-\infty}^{\infty} \phi_{xx}(\tau)\, e^{-2\pi i f \tau}\, d\tau$$

$$= \Phi_{xx}(f) \qquad (3.5.1\text{-}7)$$

3.5.2 Interpretation of the PSD as a Power Spectrum

The power in a time function x(t) is

$$\text{Total Power} = \lim_{T\to\infty} \int_{-T}^{T} |x(t)|^2\, dt \qquad (3.5.2\text{-}1)$$

$$\text{Average Power} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2\, dt \qquad (3.5.2\text{-}2)$$

Let X(f) be the Fourier transform of x(t):

$$X(f) = \int_{-\infty}^{\infty} x(t)\, e^{-2\pi i f t}\, dt \qquad (3.5.2\text{-}3)$$

The inverse Fourier transform is [34]

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(f)\, e^{2\pi i f t}\, df \qquad (3.5.2\text{-}4)$$

Plancherel's formula is [40]

$$2\pi \int_{-\infty}^{\infty} |x(t)|^2\, dt = \int_{-\infty}^{\infty} |X(f)|^2\, df \qquad (3.5.2\text{-}5)$$

Thus the total power in the time signal x(t) multiplied by 2π is the same as the total power in the Fourier transform X(f), so $|X(f)|^2$ may be interpreted as giving the power split up into frequency components, or as the power density as a function of frequency.

If x(t) is real, then X(f) = X(-f). In this case, the power over all frequencies is equal to twice the power in the positive frequencies:

$$\int_{-\infty}^{\infty} |X(f)|^2\, df = 2 \int_{0}^{\infty} |X(f)|^2\, df \qquad (3.5.2\text{-}6)$$

This fact confirms the validity of the one-sided PSD.

The PSD as the Fourier transform of the expected value of the autocorrelation function is estimated by Equation (3.5.1-3), so that

$$\Phi_{xx}(f) \cong \tilde{\Phi}_{xx}(f)$$

$$\cong \frac{1}{2T} |X(f)|^2 \qquad (3.5.2\text{-}7)$$

$\cong$ average power in x(t) broken into frequency components

3.5.3 Discrete Data

It is most common for data to be taken at discrete times using a digital computer. Assuming that the samples of the continuous process x(t) are evenly spaced in time from -T to T, Equation (3.5.1-4) is approximated by

$$X(f) \equiv \sum_{k=0}^{N-1} x[(k+\tfrac{1}{2})\Delta t - T]\, e^{-2\pi i f[(k+.5)\Delta t - T]}\, \Delta t \qquad (3.5.3\text{-}1)$$

where

$$\Delta t = \frac{2T}{N} \text{ sec} \qquad (3.5.3\text{-}2)$$

Define the discrete function

$$x_k = x[(k+\tfrac{1}{2})\Delta t - T] \;,\; k = 0, 1, \ldots, N-1 \qquad (3.5.3\text{-}3)$$

and the discrete Fourier transform

$$X_j = \sum_{k=0}^{N-1} x_k\, e^{-2\pi i j k/N} \;,\; j = 0, 1, \ldots, N-1 \qquad (3.5.3\text{-}4)$$

Equation (3.5.3-4) has a rigorous discrete inverse Fourier transform [6]

$$x_k = \frac{1}{N} \sum_{j=0}^{N-1} X_j\, e^{2\pi i j k/N} \;,\; k = 0, 1, \ldots, N-1 \qquad (3.5.3\text{-}5)$$

For j = 0, Equation (3.5.3-4) shows that $\frac{X_o}{N}$ is the average or mean of the data. For numerical reasons, the mean is often removed from the data without affecting the remaining Fourier transform coefficients, in which case the DC term becomes $X_o = 0$.

The discrete equivalent to Plancheral's theorem is Parseval's theorem [40]

$$\sum_{j=0}^{N-1} |X_j|^2 = N \sum_{k=0}^{N-1} |x_k|^2 \qquad (3.5.3\text{-}6)$$

This theorem shows that the power density interpretation of the continuous PSD holds for the discrete PSD.

Substituting Equation (3.5.3-3) into Equation (3.5.3-1) yields

$$X(f) \cong e^{-2\pi if(\Delta t/2 - T)} \sum_{k=0}^{N-1} x_k e^{-2\pi ifk\Delta t} \Delta t \qquad (3.5.3\text{-}7)$$

so that except for the phase change from the complex exponential in front of the summation sign due to the shift in the time origin (which does not affect the PSD), the discrete Fourier transform of Equation (3.5.3-4) approximates the continuous Fourier transform at discrete frequencies f_j by

$$X(f_j) \cong X_j \Delta t \qquad (3.5.3\text{-}8)$$

with

$$f_j = \frac{j}{N\Delta t} = \frac{j}{2T} \text{ Hz} \qquad (3.5.3\text{-}9)$$

Using sampled data over a time span 2T with time spacing Δt and $2T = N\Delta t$, the discrete Fourier transform is thus computed at a frequency spacing of

$$\Delta f = \frac{1}{2T} \text{ Hz} \qquad (3.5.3\text{-}10)$$

If x(t) for $-\infty < t < \infty$ is a band-limited signal such that $X(f) = 0$ for $|f| > F$, and X_j is the discrete Fourier transform of samples of x(t) at a sampling interval Δt with

$$N\,\Delta f = \frac{N}{2T} = \frac{N}{N\Delta t} = \frac{1}{\Delta t} \geq 2F$$

then x(t) can be exactly reconstructed from X_j [11]. Thus the finite Fourier transform given by X_j will contain all of the magnitude information about the time series x(t) only if the sampling frequency is greater than 2F. This is the Nyquist criterion. If the time series x(t) is not band-limited, then the lower frequency discrete Fourier transform values will reflect energy from the higher frequency components of the signal. This phenomenon is known as aliasing [6].

If the data x_k are real, Equation (3.5.3-4) shows that

$$X_{N-j} = X_j^* , \quad j = 1, 2, \ldots, \left[\frac{N}{2}\right] \qquad (3.5.3\text{-}11)$$

where $\left[\frac{N}{2}\right]$ is the largest integer less than or equal to $\frac{N}{2}$. This dependence is due to the Nyquist criterion. Information is only available about the power content of the signal at frequencies up to $\frac{1}{2\Delta t}$.

The fact that the discrete Fourier transform has been defined for positive frequencies and the Nyquist criterion combine to limit the range of available frequency information to $0 \le f \le \frac{1}{2\Delta t}$. Thus by Equations (3.5.1-5) and (3.5.3-8), the estimate from a finite span of sampled data of the one-sided PSD at frequency f_j is

$$\tilde{\Phi}^1_{xx}(f_j) = \frac{\Delta t^2}{T} |X_j|^2 \ , \ j = 0, 1, ..., \left[\frac{N}{2}\right] - 1 \qquad (3.5.3\text{-}12)$$

$$\tilde{\Phi}^1_{xx}(f_{[N/2]}) = \begin{cases} \frac{\Delta t^2}{T} |X_{[N/2]}|^2 & \text{if N is odd} \\ \frac{\Delta t^2}{2T} |X_{[N/2]}|^2 & \text{if N is even} \end{cases} \qquad (3.5.3\text{-}13)$$

Equation (3.5.3-13) guarantees that the one-sided PSD maintains the symmetry required by the Nyquist criterion.

The discrete Fourier transform can be evaluated using the Cooley-Tukey Fast Fourier Transform (FFT) algorithm if N is a power of 2 [6]. This well known algorithm greatly reduces the computational burden of calculating the discrete Fourier transform, with the number of computations being proportional to $N \log_2 N$ rather than the N^2 required for the brute force evaluation of Equation (3.5.3-4) [6].

3.6 Frequency Averaging

In Section 3.5.1, it was shown that the estimate of the power spectral density is unbiased. However, it can also be shown that for most processes that are of interest, the standard deviation of this estimate at a given frequency is equal to the actual value of the PSD at that frequency [6], [17], [33].

This makes Equation (3.5.1-3) a poor estimate of $\Phi_{xx}(f)$. The standard deviation of the error is not a function of the time interval of available data [6], [17], [33]. This means that no matter how large T becomes, the standard deviation will not improve. The variance of the estimate may be decreased by either averaging several estimates or averaging adjacent frequencies of a given estimate. The former method is known as ensemble averaging, while the latter is called frequency averaging.

Frequency averaging will be employed in this thesis because a collection of sample paths is not always available for processing. This method is valid because the errors in the estimates of adjacent points on the PSD are nearly independent [17]. Frequency averaging has the disadvantage of decreasing the resolution of the PSD because several points on the graph are replaced by one. However, because a PSD is commonly plotted using a log-log scale, this loss of detail may be utilized to eliminate the "bunching up" of points at the high frequency end of the graph where the human eye typically cannot resolve individual frequencies. This is done by averaging fewer points at the lower frequencies and more at the higher end. This process results in a PSD that is more accurate at the higher frequencies because more points are involved in averaging, but which still retains good resolution at lower frequencies.

The averaging system used in this thesis is described by the Table 3.4-1 [22]. The first 32 PSD data points are not averaged. Starting with the 33rd point, the PSD data is divided into sets of points with indices that range from $2^{k-1} + 1$ to 2^k. The points within each range are then averaged to leave only 16 points. This system reduces the number of points in the PSD to approximately 50 points per decade.

Table 3.4-1 Logarithmic Frequency Averaging of 32,768 Points

Data Interval	Points in Interval	Number of Points per Averaging	Number of Points to Plot
1 - 32	32	1	32
31 - 64	32	2	16
65 - 128	64	4	16
129 - 256	128	8	16
257 - 512	256	16	16
513 - 1024	512	32	16
1025 - 2048	1024	64	16
2049 - 4096	2048	128	16
4097 - 8192	4096	256	16
8193 - 16384	8192	512	16
16385 - 32768	16384	1024	16

Chapter 4

Markov Stochastic Processes

4.1 Martingale and Markov Processes

A martingale is a stochastic process $x(t,\omega)$ for which [13]

$$E\{\underline{x}(t_2) \mid \underline{x}(t_1)\} = \underline{x}(t_1) \quad \text{for } t_2 > t_1 \tag{4.1-1}$$

where $E\{\bullet \mid \bullet\}$ denotes conditional expectation. In other words, for a martingale the expected value of the future given the present is equal to the present. The name martingale is the French term for the gambling strategy of redoubling a bet until a win is obtained. This eventually has to occur in a fair game, except that the house always has the odds slightly in its favor, and except for the theorem of gamblers ruin. This states that in a fair game a gambler with a finite stake betting against a house with an infinite stake will eventually lose his fortune with probability 1 [12].

A Markov stochastic process $\underline{x}(t,\omega)$, as defined by Markov around 1906 [2], is one for which [18]

$$E\{\underline{x}(t_{n+1}) \mid \underline{x}(t_n), \underline{x}(t_{n-1}), \ldots, \underline{x}(t_1)\} = E\{x(t_{n+1}) \mid x(t_n)\} \tag{4.1-2}$$

In other words, for a Markov process the future depends on the present and not on the complete past history of the process. This property is similar to the principle of causality in physics, where differential equations predicting the future behavior of a system depend only on the initial conditions of the system independently of how the system reached those initial conditions.

Presuming the validity of the principle of causality, one would expect to be able to model the noise processes in physical systems as Markov processes. However, if a system is modeled with an incomplete set of states, then the initial conditions of the reduced set of states do not completely specify the future. Hence an incomplete set of noise states could lead to what seems to be a non-Markov process.

Chapter 5 defines fractional Brownian motion as an example of a non-Markov process. In-so-far as the log-log PSD slopes in fractional Brownian motion are seen in experimental data, one could suppose that the true underlying physical process could be modeled with Markov noise states, if enough of them are included. However, it might be inconvenient to include a sufficient number of states to model a distributed parameter system.

The remainder of this chapter discusses Markov processes, in particular the Brownian motion process and the processes derived from it.

4.2 Weiner Brownian Motion Process

A Weiner process or Brownian motion process or random walk process $\beta(t,\omega)$ is a stochastic process such that [18]

(1) $\beta(t)$ has stationary independent increments, that is if $t_1 < t_2 < t_3 < t_4$, then $\beta(t_2) - \beta(t_1)$ is independent of $\beta(t_4) - \beta(t_3)$, and the probability distribution of $\beta(t_2) - \beta(t_1)$ only depends on $t_2 - t_1$ and not on t_1 and t_2 individually.

(2) For every t, $\beta(t)$ is normally distributed with zero mean. Hence the increment $\beta(t_2) - \beta(t_1)$ is normally distributed with zero mean, and by (1) the variance of the increment $\beta(t_2) - \beta(t_1)$ depends only on $t_2 - t_1$.

(3) One can specify that $Pr[\beta(0)=0] = 1$, and only consider $\beta(t)$ for $t \geq 0$. However, one can also consider $\beta(t)$ for $-\infty < t < \infty$.

A stochastic process with these characteristic properties can be mathematically constructed [20], so the Weiner process exists. It is a martingale, because for $t_2 > t_1$

$$\begin{aligned} E\{\beta(t_2) \mid \beta(t_1)\} &= E\{ [\beta(t_2)-\beta(t_1)] + [\beta(t_1) - \beta(0)] \mid [\beta(t_1) - \beta(0)] \} \\ &= E\{ [\beta(t_1) - \beta(0)] \mid [\beta(t_1) - \beta(0)] \} \\ &= \beta(t_1) \end{aligned} \qquad (4.2\text{-}1)$$

by the stationary independent increment property and assuming $\beta(0) = 0$.

Among the properties of a Weiner process that can be derived from the defining properties are that any given sample path is almost surely (with probability 1) continuous and almost surely non-differentiable [18].

The Weiner process models the behavior of the Brownian motion of a small particle suspended in a fluid and continually buffeted by collisions with fluid molecules, as first observed for pollen particles in water by the biologist Brown in 1828. Einstein worked out some of the statistical mechanical characteristics of Brownian motion in 1905, Weiner put it on a firm mathematical footing in the 1920s, and Levy further investigated its properties in the 1930s and 1940s. [19]

There are trillions of molecular collisions per second with a particle undergoing Brownian motion. In between collisions, the particle motion is differentiable in classical physics. Because the collision of a molecule with a particle actually only involves the interaction of force fields, the motion through a collision can also be regarded as being differentiable. Therefore, a physical Brownian motion sample path is everywhere differentiable, whereas the stochastic process which models it very well and has useful properties, e.g., for defining stochastic integrals, is nowhere differentiable.

Let $v(t_2-t_1)$ be the variance of $\beta(t_2) - \beta(t_1)$. By the stationarity and independence of the increments,

$$\begin{aligned} v(2\tau) &= E\{ [\beta(t+2\tau) - \beta(t)]^2 \} \\ &= E\{ [(\beta(t+2\tau) - \beta(t+\tau)) + (\beta(t+\tau) - \beta(t))]^2 \} \end{aligned}$$

$$= E\{ [\beta(t+2\tau) - \beta(t+\tau)]^2 \} + E\{ [\beta(t+\tau) - \beta(t)]^2 \}$$

$$= 2\, v(\tau) \qquad (4.2\text{-}2)$$

Similarly $v(n\tau) = nv(\tau)$ and $v(\tau/n) = v(\tau)/n$ for any integer n, so that $v(a\tau) = av(\tau)$ for any rational number a, or any real number a if $v(\tau)$ is assumed continuous. Thus $v(\tau) = v(1\cdot\tau) = v(1)\tau$, so that if $v(\tau)$ is continuous

$$v(\tau) = E\{ [\beta(t_2) - \beta(t_1)]^2 \} = \sigma^2 \, | t_2 - t_1 | \qquad (4.2\text{-}3)$$

where σ is called the standard deviation and σ^2 the variance of the Weiner process [9], [35].

The autocorrelation function of a Weiner process for $t_1 < t_2$ assuming $\beta(0) = 0$ is

$$\phi_{\beta\beta}(t_1,t_2) = E\{\beta(t_1)\beta(t_2)\}$$

$$= E\{ \beta(t_1)[\beta(t_1) + (\beta(t_2) - \beta(t_1))] \}$$

$$= E\{ [\beta(t_1) - \beta(0)][(\beta(t_1) - \beta(0)) + (\beta(t_2) - \beta(t_1)] \}$$

$$= \sigma^2 t_1 \qquad (4.2\text{-}4)$$

so that

$$\phi_{\beta\beta}(t_1,t_2) = \sigma^2 \min(t_1,t_2) \qquad (4.2\text{-}5)$$

4.3 White Noise PSD Slope

White noise is a zero mean, Gaussian, stationary, stochastic process $x(t,\omega)$ which has a delta function for its autocorrelation function. The Fourier transform of a unit delta function is 1 for all frequencies. Thus the two-sided PSD of white noise has a constant level of σ^2/Hz for all frequencies. This stochastic process is called white noise in analogy with white light, which has more or less equal energy at all visible frequencies. The log-log PSD slope of white noise is 0.

Having a constant PSD level for all frequencies is a physical impossibility, because it implies infinite power. However, the concept of white noise is mathematically useful for modeling physical systems. This is done in two ways.

First, for sampled observations, measurement noise is often assumed to be independent from one discrete sample time to the next. This whiteness assumption causes no physical difficulty because of the discrete sample times.

Second, white noise is introduced as a driving force in state differential equations through the perfectly rigorous definition of the stochastic integral with respect to the differential (rather than derivative) of a Weiner process. The state dynamics then shape the white driving noise to give a colored noise response. Here, the term colored noise refers to a stochastic process which has different energy content (or PSD level) at different frequencies, in analogy with colored light.

4.4 PSD Slope of Random Walk, Trend, and Quantization Noise

A Weiner or random walk process has stationary independent increments, but itself is not stationary. Hence its PSD is not defined in the sense of the Fourier transform of a stationary autocorrelation function. However, a PSD can be computed from the magnitude squared of the Fourier transform of a finite segment of a random walk sample path.

Either by considering approximations to a Weiner process or by doing formal manipulations with its correlation function, it can be shown that both the approximate and formal derivative of a Weiner process is white noise [18]. Thus, the PSD of the Weiner process has a -2 log-log slope. This can be seen by using integration by parts in the Fourier transform of the derivative of the Weiner process g(t) with $g(-\infty) = g(\infty) = 0$,

$$\int_{-\infty}^{\infty} \frac{dg(t)}{dt} e^{-2\pi i f t}\, dt = 2\pi i f \int_{-\infty}^{\infty} g(t)\, e^{-2\pi i f t}\, dt \qquad (4.4\text{-}1)$$

Thus the Fourier transform of random walk is equal to $\frac{1}{2\pi if}$ times that of white noise. Using Equation (3.5.1-3) to arrive at the PSD shows that the PSD of random walk is equal to $\frac{1}{4\pi^2 f^2}$ multiplied by the PSD of white noise. This has the effect of decreasing the log-log slope of the PSD by two. Because Equation (4.4-1) holds for any function g(t), integrating any function will decrease its log-log PSD slope by two, and equivalently differentiating a function will increase its PSD slope by the same amount.

Now consider the Fourier transform of any finite segment of a trend αt, $-T \le t \le T$:

$$X(f) = \int_{-T}^{T} \alpha t \, e^{-2\pi ift} \, dt \qquad (4.4\text{-}2)$$

Equations (3.5.3-2) and (3.5.3-9) motivate the change of variables

$$f = j\frac{1}{2T}, \quad t = n\Delta t = n\frac{2T}{N} \qquad (4.4\text{-}3)$$

where, for the moment, j and n are allowed to vary continuously. This leads to

$$X(j) = \int_{-N/2}^{N/2} \alpha \, n\Delta t \, e^{-2\pi ijn/N} \, \Delta t \, dn \qquad (4.4\text{-}4)$$

Using integration by parts Equation (4.4-4) becomes

$$X(j) = \alpha \Delta t^2 \left[\frac{-N^2}{4\pi ij} \left(e^{-\pi ij} + e^{\pi ij} \right) + \frac{N^2}{4\pi^2 j^2} \left(e^{-\pi ij} - e^{\pi ij} \right) \right] \qquad (4.4\text{-}5)$$

If, at this point, j is restricted to take on only integer values as it would in the discrete Fourier transform, this equation reduces to

$$X(j) = \frac{\alpha \Delta t^2 N^2}{2\pi j} (-1)^j i$$

$$= \frac{2\alpha T^2}{\pi j} (-1)^j i \qquad (4.4\text{-}6)$$

Therefore, by Equation (3.5.1-3) the value of the PSD at frequencies spaced $\frac{1}{2T}$ Hz apart is given by

$$\Phi_{xx}(f) = \frac{2\alpha^2 T^3}{\pi^2 f^2} \tag{4.4-7}$$

which will have a -2 log-log slope.

Because both random walk and trend have a -2 log-log PSD slope, they cannot be distinguished using frequency domain analysis. This is one motivation for developing the maximum likelihood estimator of Chapter 6 which is capable of separating the trend and random walk.

Quantization noise results from discretization of continuous signals. The PSD of quantization noise is derived in this section because it is present in the experimental data presented in Section 6.8.1.

If a continuous signal is being measured by a digital device, the true value of the signal will be truncated to the number of decimal places stored by the instrument. Under the assumption that the value of the signal is changing rapidly with respect to the level of quantization, this is equivalent to subtracting a uniformly distributed random variable n(t) from the true value of the signal. The probability distribution of n is given by

$$p[n(t)] = \begin{cases} \frac{1}{q} & \text{if } 0 \le n \le q \\ 0 & \text{else} \end{cases} \tag{4.4-8}$$

where q is the resolution of the measurement device. Quantization noise is assumed to be uncorrelated in time so that

$$\begin{aligned} \phi_{nn}(\tau) &= E\{n(t+\tau)n(t)\} \\ &= E\{n(t)^2\}\,\delta(\tau) \\ &= \frac{q^2}{3}\,\delta(\tau) \end{aligned} \tag{4.4-9}$$

where $\delta(\tau)$ is the delta function.

The PSD of quantization noise is

$$\Phi_{nn}(f) = \int_{-\infty}^{\infty} \phi_{nn}(\tau)\, e^{-2\pi i f \tau}\, d\tau$$

$$= \frac{q^2}{3} \qquad (4.4\text{-}10)$$

As expected, this PSD is a constant because the quantization noise is assumed to be uncorrelated in time.

If a measurement y(t) is made up of quantization noise subtracted from a stationary signal x(t), then its autocorrelation function is given by

$$\phi_{yy}(\tau) = E\{ y(t+\tau)y(t) \}$$

$$= E\{ [x(t+\tau) - n(t+\tau)]\, [x(t) - n(t)] \}$$

$$= E\{ x(t+\tau)x(t) \} + E\{ n(t+\tau)n(t) \}$$

$$= \phi_{xx}(\tau) + \phi_{nn}(\tau) \qquad (4.4\text{-}11)$$

if the x(t) and n(t) are independent.

The PSD of y(t) is then given by

$$\Phi_{yy}(f) = \int_{-\infty}^{\infty} [\phi_{xx}(\tau) + \phi_{nn}(\tau)]\, e^{-2\pi i f \tau}\, d\tau$$

$$= \Phi_{xx}(f) + \Phi_{nn}(f) \qquad (4.4\text{-}12)$$

As a result of Equation (4.4-12), PSDs which are calculated from experimental data will show a flattening to zero at the higher frequencies where quantization noise becomes dominant unless a low-pass filter is used to remove it.

In the experimental data used in Chapter 6, quantization noise shows up with a +2 log-log slope. This is because the PSD was produced using the difference of successive samples that were corrupted by quantization noise. Recall that Equation (4.4-1) shows that differentiation of the signal of interest increases the log-log PSD slope by two, and that the differencing operation is essentially differentiation.

4.5 Ito Stochastic Integral

If $g(t,\omega)$ is a random function, and $\beta(t,\omega)$ is a Weiner process such that $g(s,\omega)$ is independent of $\beta(t,\omega)$ for $s \le t$, then the Ito stochastic integral of g with respect to $d\beta$ for given sample paths is defined to be [18]

$$\int_a^b g(t,\omega)\, d\beta(t,\omega) = \lim_{\text{mesh}\to 0} \sum_{j=1}^{n-1} g(t_j,\omega)\,[\beta(t_{j+1},\omega) - \beta(t_j,\omega)] \qquad (4.5\text{-}1)$$

where $a = t_1 < t_2 < \ldots < t_n = b$ is a partition of the interval [a,b] with mesh = max $(t_{j+1}-t_j)$. In this definition, it is imprecise in what sense the limit is taken. In Reference [18] it is taken in the sense of limit in the mean (l.i.m.) for the mean square integral calculus. The completely rigorous definition utilizes the machinery of Lebesgue integration with limits in probability [30]. Note that the differential notation $d\beta$ does not involve taking a derivative, and has a rigorous meaning in the definition of the stochastic integral.

The Stratonovich stochastic integral has the function g evaluated at the midpoint of each sub-interval in the partition, instead of the left end point used in the Ito definition. Formal properties of calculus result for, e.g., changes of integration variable. However the Ito definition is preferred, even though formal calculus formulas are lost, because it is defined on a larger class of functions and good probability properties result, including the fact that the Ito stochastic integral is a martingale. [2]

The Ito and Stratonovich stochastic integrals are the same for a non-random function g(t) and thus both may be evaluated using the formal rules of calculus as if $\beta(t)$ were continuously differentiable [18]. Weiner had defined the stochastic integral in this case [10]. The Ito and Stratonovich integrals give different results for a random function $g(t,\omega)$, in particular different expected values over the ensemble of all sample paths [2].

4.6 Stochastic Differential Equations

The stochastic differential equation for a vector state $\underline{x}$ and vector Weiner process $\underline{\beta}$ is written in terms of differentials as

$$d\underline{x} = F(\underline{x},t)\,dt + G(\underline{x},t)\,d\underline{\beta}(t) \qquad (4.6\text{-}1)$$

This notation stands for the Ito stochastic integral equation

$$\underline{x}(b) - \underline{x}(a) = \int_a^b F(\underline{x},t)\,dt + \int_a^b G(\underline{x},t)\,d\underline{\beta}(t) \qquad (4.6\text{-}2)$$

If the functions F and G satisfy Lipschitz and other conditions, and if the random variable $\underline{x}(a)$ with $E[\underline{x}(a)^2] < \infty$ is independent of $\underline{\beta}(t)$ for $t \geq a$, then there is a unique Markov stochastic process solution $\underline{x}(t)$ of Equation (4.6-2) in the mean square sense such that $\underline{x}(t) - \underline{x}(a)$ is independent of $\underline{\beta}(\tau)$ for $\tau \geq t$ [18].

This result depends on the stationary independent increment property of the scalar process $\beta(t)$ with $E\{ [\beta(t_2) - \beta(t_1)]^2 \} = \sigma | t_2 - t_1 |$, where σ is the standard deviation of the increments. The stochastic integral cannot be defined with fractional Brownian motion β_H (see Chapter 5) replacing β, since its increments are not independent, even though they are stationary, and $E\{ [\beta_H(t_2) - \beta_H(t_1)]^2 \} = \sigma | t_2 - t_1 |^{2H}$.

4.7 Kolmogorov Fokker-Planck Partial Differential Equations and Derivation of the Kalman Filter

The propagation of the probability density of the initial condition $\underline{x}(a,\omega)$ of Equation (4.6-2) into the probability density of the solution $\underline{x}(t,\omega)$ can be shown to satisfy a parabolic partial differential equation called the Kolmogorov Fokker-Planck equation [18]. If the state equation (4.6-1) or (4.6-2) is linear and the initial condition Gaussian, then the rigorous formulas for the propagation of the resulting Gaussian probabilty density lead to equations for the propagation of the Gaussian mean and covariance, and then to Kalman filter formulas for the Bayes estimation of the mean and its covariance conditioned on observations [18].

4.8 Exponentially Correlated Noise

Consider the stochastic differential equation

$$dx = -c_1 x(t)\,dt + c_1 c_2\,d\beta(t) \tag{4.8-1}$$

As noted in Section 4.5, because the coefficients of $d\beta(t)$ are nonrandom, this equation may be evaluated using the formal rules of calculus. Thus the equation reduces to

$$\frac{dx}{dt} = -c_1 x(t) + c_1 c_2 w(t)\ ,\ x(0) \sim N(0, \frac{c_2^2 c_1}{2}) \tag{4.8-2}$$

so that [18]

$$x(t) = x(0)\,e^{-c_1 t} + c_1 c_2 \int_0^t e^{-c_1(t-s)}\,w(s)\,ds \tag{4.8-3}$$

where w(t) is a zero mean, white, Gaussian forcing function. The variance of the initial condition is chosen for convenience in deriving the autocorrelation function. Because x(0) and w(t) have mean zero, x(t) also has mean zero. In this thesis, c_1 will be referred to as the inverse or reciprocal time constant and c_2 as the scaling parameter.

The autocorrelation function for x(t) assuming $t_1 \geq t_2$ is [18]

$$\phi_{xx}(t_1,t_2) = E\{x(t_1)x(t_2)\}$$

$$= E\left\{\left[x(0)\,e^{-c_1 t_1} + c_1 c_2 \int_0^{t_1} e^{-c_1(t_1-s)}\,w(s)\,ds\right] \times \left[x(0)\,e^{-c_1 t_2} + c_1 c_2 \int_0^{t_2} e^{-c_1(t_2-\xi)}\,w(\xi)\,d\xi\right]\right\}$$

$$= c_2^2\,(c_1/2)\,e^{-c_1(t_1+t_2)} + c_1^2 c_2^2 \int_0^{t_1}\int_0^{t_2} e^{-c_1(t_1-s)}\,e^{-c_1(t_2-\xi)}\,\delta(s-\xi)\,ds\,d\xi$$

$$= c_2^2 (c_1/2)\, e^{-c_1(t_1+t_2)} + c_1^2 c_2^2 \int_0^{t_2} e^{-c_1(t_1-\xi)}\, e^{-c_1(t_2-\xi)}\, d\xi$$

$$= c_2^2 (c_1/2)\, e^{-c_1(t_1-t_2)} \quad , \; t_1 \geq t_2 \tag{4.8-4}$$

The fact that the correlation function for x(t) is an exponential has lead to its being called exponentially correlated noise.

Because exponentially correlated noise is stationary, its PSD may be calculated directly as the Fourier transform of its autocorrelation function:

$$\Phi_{xx}(f) = \int_{-\infty}^{\infty} c_2^2 (c_1/2)\, e^{-c_1|\tau|}\, e^{-2\pi i f \tau}\, d\tau$$

$$= \frac{1}{2\pi} \frac{(c_1 c_2)^2}{f^2 + (c_1/2\pi)^2} \tag{4.8-5}$$

where $\tau = t_1 - t_2$. This function will have a log-log slope of zero for frequencies below $(c_1/2\pi)$, near this frequency it will begin to transition to a -2 log-log slope. Because exponentially correlated noise has a PSD which shows more energy in some frequencies than others, it is often referred to as colored noise.

4.9 Simulation of Markov Noise Processes

The Markov noise processes that were described in the preceding sections may all be simulated using a digital computer. Computer generated sample paths will be useful for evaluating the performance of the algorithms presented in Chapters 6 and 7. Subroutines that produce independent, zero mean, unit variance, normally distributed random variables, and others that return independent, uniformly distributed random variables are commonly available in software libraries. These subroutines make use of random number generation algorithms.

White noise is a sequence of normally distributed random variables, while random walk and exponentially correlated noise are simple functions of a white noise sequence. Quantization noise may be simulated using a random number generator with a uniform distribution or by truncating data to a given precision.

A zero mean white sequence is formed using the following formula:

$$y(t_i) = \sigma\, w(t_i) \tag{4.9-1}$$

where σ is the desired standard deviation of the sequence and $w(t_i)$ is a zero mean, unit variance, Gaussian random variable from a random number generator.

A random walk sample path may be simulated by

$$y(t_i) = y(t_{i-1}) + b\,\Delta t^{1/2}\, w(t_i) \tag{4.9-2}$$

where b is the random walk standard deviation parameter and Δt is the desired time step.

Exponentially correlated noise is generated using the discretized form of Equation (4.8-2):

$$y(t_i) = e^{-c_1 \Delta t}\, y(t_{i-1}) + c_1 c_2\, \Delta t^{1/2}\, w(t_i) \tag{4.9-3}$$

This process may also be simulated using a first order approximation to Equation (4.9-3):

$$y(t_i) = y(t_{i-1}) - c_1 y(t_{i-1})\Delta t + c_1 c_2 \Delta t^{1/2}\, w(t_i) \tag{4.9-4}$$

Chapter 5

Fractional Brownian Motion as an Example of a Non-Markov Stochastic Process

5.1 Motivation for Defining Fractional Brownian Motion

Fractional Brownian motion (fBm) is a generalization of the more familiar Brownian motion (Weiner) process that was defined in Section 4.2. The definition of fBm was motivated by the fact that there are many signals which occur in physical systems that have PSDs with log-log slopes between 0 and -2 over a very wide range of frequencies. Such processes cannot be modeled by a combination of a reasonable number of Markov processes.

A PSD which is proportional to f^{β} over all frequencies for $-2 < \beta < 0$ is indicative of a long term dependence in the data [26]. This means that the value of the process at the current time is highly dependent upon all past values of the process. This is why Markov processes cannot model systems of this type, because the future value of a Markov process is only dependent upon its current value.

Some examples of systems which exhibit a power spectral density proportional to f^{β} are semiconductors [36], loudness and pitch fluctuations in music and speech [43], seismic reflections [42], sunspot activity [27], and the classic example of Nile river flooding [27]. While the physical processes of these systems are different, fBm provides a common mathematical model which does characterize them in a useful way.

5.2 Definition of Fractional Brownian Motion

The definition of fBm was proposed by Mandlebrot and van Ness as a modification of an earlier definition by Barnes and Allen [4]. An fBm process is defined by the following equations [26]:

$$\beta_H(0,\omega) = b_0$$

$$\beta_H(t,\omega) - \beta_H(0,\omega) = \frac{1}{\Gamma(H+.5)}\left\{\int_{-\infty}^{0}\left[(t-s)^{H-1/2} - (-s)^{H-1/2}\right]d\beta(s,\omega) + \int_{0}^{t}(t-s)^{H-1/2}\,d\beta(s,\omega)\right\} \quad (5.2\text{-}1)$$

In this definition, $\beta(t,\omega)$ is a unit variance Brownian motion or Weiner process at time t with probability space ω, $d\beta(t,\omega)$ is its differential for the Ito stochastic integral, and $\Gamma(\bullet)$ is the gamma function. The parameter H is the fBm parameter which will define the slope of the fBm PSD. As Section 5.6.2 will show, the log-log slope of the PSD will be equal to -(1+2H). The definition is valid for $0 < H < 1$. In this thesis, it will be assumed that $b_0 = 0$. The fact that the integral in Equation (5.2-1) is evaluated from minus infinity to the current time makes fBm a non-Markov process, except in the case where $H = \frac{1}{2}$ where $\beta_{1/2}$ is ordinary Brownian motion:

$$\beta_{1/2}(t,\omega) = \int_{0}^{t} d\beta(s,\omega) \quad (5.2\text{-}2)$$

Because fBm is based upon Brownian motion, it is a zero mean Gaussian process.

Equation (5.2-1) is actually the fractional integral [32] of the white noise process $d\beta(t,\omega)$ which would more commonly be written as [26]

$$\beta_H(t_2,\omega) - \beta_H(t_1,\omega) = \frac{1}{\Gamma(H+.5)}\left\{ \int_{-\infty}^{t_2} \left[(t_2-s)^{H-1/2}\right] d\beta(s,\omega) - \int_{-\infty}^{t_1} (t_1-s)^{H-1/2}\, d\beta(s,\omega) \right\} \quad (5.2\text{-}3)$$

except that the individual integrals in Equation (5.2-3) are divergent, in contrast with the combined integrals in Equation (5.2-1).

5.3 Self-similarity of fBm

A useful property of fBm is that its increments are self-similar. Self-similarity means that

$$p[\beta_H(t+a\Delta t) - \beta_H(t)] = p[\Delta t^H \beta_H(a)] \quad (5.3\text{-}1)$$

where $p(\bullet)$ indicates a probability density [26]. Note that this equation also implies that the fBm increments are stationary, though they are only independent when $H = \frac{1}{2}$.

Without loss of generality, Equation (5.3-1) can be shown to be true for $b_o = 0$ and $t = 0$. Using Equation (5.2-1) and dropping the explicit dependence on the probability space ω for simplicity

$$\beta_H(a\Delta t) = \frac{1}{\Gamma(H+.5)}\left\{ \int_{-\infty}^{0} \left[(a\Delta t-s)^{H-1/2} - (-s)^{H-1/2}\right] d\beta(s) + \int_{0}^{a\Delta t} (a\Delta t-s)^{H-1/2}\, d\beta(s) \right\} \quad (5.3\text{-}2)$$

The change of variable $s' = \frac{s}{\Delta t}$ leads to

$$d\beta(s') = \Delta t^{1/2}\, d\beta(s) \quad (5.3\text{-}3)$$

$$\beta_H(a\Delta t) = \Delta t^H \frac{1}{\Gamma(H+.5)} \left\{ \int_{-\infty}^{0} \left[(a-s')^{H-1/2} - (-s')^{H-1/2} \right] d\beta(s') + \int_{0}^{a} (a-s')^{H-1/2}\, d\beta(s') \right\}$$

$$= \Delta t^H \beta_H(a) \qquad (5.3\text{-}4)$$

5.4 Autocorrelation of fBm

In order to find the autocorrelation function of fBm it is first useful to find the variance of an fBm increment. By self-similarity

$$p[\beta_H(t+\Delta t) - \beta_H(t)] = p[\Delta t^H \beta_H(1)] \qquad (5.4\text{-}1)$$

Thus in the case where the Brownian motion process $\beta(t)$ of Equation (5.2-1) has unit variance, the variance of an fBm increment is given by

$$\begin{aligned} E\{ [\beta_H(t+\Delta t) - \beta_H(t)]^2 \} &= E\{ [\Delta t^H \beta_H(1)]^2 \} \\ &= \Delta t^{2H} E\{\beta_H(1)^2\} \\ &= \Delta t^{2H} V_H \end{aligned} \qquad (5.4\text{-}2)$$

where

$$V_H \equiv E\{\beta_H(1)^2\} \qquad (5.4\text{-}3)$$

The parameter V_H is only a function of H, so it would seem that H is the only parameter required to characterize fBm. However, it is possible to create a more "noisy" sample path by multiplying the sample path of Equation (5.2-1) by a positive constant A. As can be seen by examining Equation (5.4-2) this would have the effect of scaling the variance of an increment by A^2. Thus a new parameter will be introduced to account for possible scaling of the fBm process:

$$\sigma_H^2 \equiv A^2 V_H \qquad (5.4\text{-}4)$$

This parameter is introduced because it is common practice ([4], [24], [26]) to consider V_H to describe the unscaled process. In this thesis, the parameters H and σ_H will be used to completely characterize an fBm process.

Assuming that $\beta_H(0) = 0$, so that $E\{\beta_H(0)\beta_H(t_i)\} = 0$, the autocorrelation function is now given by

$$\begin{aligned}
\phi_{\beta\beta}(t_1,t_2) &= E\{\beta_H(t_1)\beta_H(t_2)\} \\
&= E\{ -\beta_H(0)\beta_H(t_1) - \beta_H(0)\beta_H(t_2) + \beta_H(t_1)\beta_H(t_2) \} \\
&= \frac{1}{2} E\{ \beta_H(0)^2 - 2\beta_H(0)\beta_H(t_1) + \beta_H(t_1)^2 \\
&\qquad + \beta_H(0)^2 - 2\beta_H(0)\beta_H(t_2) + \beta_H(t_2)^2 \\
&\qquad - \beta_H(t_1)^2 + 2\beta_H(t_1)\beta_H(t_2) - \beta_H(t_2)^2 \} \\
&= \frac{1}{2} E\{ [\beta_H(t_1) - \beta_H(0)]^2 + [\beta_H(t_2) - \beta_H(0)]^2 - [\beta_H(t_2) - \beta_H(t_1)]^2 \} \\
&= \frac{1}{2} [\, |t_1|^{2H} \sigma_H^2 + |t_2|^{2H} \sigma_H^2 - |t_2 - t_1|^{2H} \sigma_H^2 \,] \\
&= \frac{1}{2} \sigma_H^2 [\, |t_1|^{2H} + |t_2|^{2H} - |t_2 - t_1|^{2H} \,] \qquad (5.4\text{-}5)
\end{aligned}$$

When $H = \frac{1}{2}$,

$$\phi_{\beta\beta}(t_1,t_2) = \sigma_H^2 \min(t_1,t_2)$$

which is the autocorrelation function of ordinary Brownian motion.

Note also that $\phi_{\beta\beta}$ is a function of the absolute times t_1 and t_2. This makes fractional Brownian motion a nonstationary process. It is also useful to note that because fBm is a zero mean process its covariance is equal to its autocorrelation.

5.5 Autocorrelation of the Increments

It will be useful to know the autocorrelation function of the increments of an fBm process. The increments themselves are often referred to as discrete fractional Gaussian noise as a generalization of the Gaussian noise increments which comprise ordinary Brownian motion.

Let the fBm increment $X(t_k)$ at time $t_k = k\Delta t$ be defined to be

$$X(t_k) \equiv \beta_H(t_{k+1}) - \beta_H(t_k) \tag{5.5-1}$$

then the correlation between two increments is given by

$$\begin{aligned}
\phi_{xx}(t_{k+n},t_k) &= E\{ X(t_{k+n})X(t_k) \} \\
&= E\{ [\beta_H(t_{k+n+1}) - \beta_H(t_{k+n})] [\beta_H(t_{k+1}) - \beta_H(t_k)] \} \\
&= E\{ \beta_H(t_{k+n+1})\beta_H(t_{k+1}) - \beta_H(t_{k+n+1})\beta_H(t_k) \\
&\quad - \beta_H(t_{k+n})\beta_H(t_{k+1}) + \beta_H(t_{k+n})\beta_H(t_k) \} \\
&= -\tfrac{1}{2} E\{ [\beta_H(t_{k+n+1}) - \beta_H(t_{k+1})]^2 \} + \tfrac{1}{2} E\{ [\beta_H(t_{k+n+1}) - \beta_H(t_k)]^2 \} \\
&\quad + \tfrac{1}{2} E\{ [\beta_H(t_{k+n}) - \beta_H(t_{k+1})]^2 \} - \tfrac{1}{2} E\{ [\beta_H(t_{k+n}) - \beta_H(t_k)]^2 \} \\
&= -\tfrac{1}{2} (|n|\Delta t)^{2H} \sigma_H^2 + \tfrac{1}{2} (|n+1|\Delta t)^{2H} \sigma_H^2 \\
&\quad + \tfrac{1}{2} (|n-1|\Delta t)^{2H} \sigma_H^2 - \tfrac{1}{2} (|n|\Delta t)^{2H} \sigma_H^2 \quad \text{by Equation (5.4-2)} \\
&= \tfrac{1}{2} \Delta t^{2H} \sigma_H^2 [\, |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \,]
\end{aligned} \tag{5.5-2}$$

Note that $\phi_{xx}(t_{k+n},t_k) = \phi_{xx}(n)$ so that the increments are correlation stationary. Also, because the increments are zero mean, the correlation function is equal to the covariance of the increments just as in the case of the fBm process itself.

5.6 fBm Derivative Process

It is possible to define a derivative process for fBm as follows [26]:

$$\beta_{H,\delta}(t) = \frac{1}{\delta}[\beta_H(t+\delta) - \beta_H(t)] \qquad (5.6\text{-}1)$$

In the limit as δ approaches zero, this process does not exist because fBm is not differentiable [26]. However, this process is an approximation to the derivative of fBm which will allow the calculation of the autocorrelation function and PSD of fractional Gaussian noise.

5.6.1 Autocorrelation of fBm Derivative Process

The autocorrelation function of Equation (5.6-1) is given by

$$\phi_{\beta\beta}(t_2,t_1) = \frac{1}{\delta^2} E\{ [\beta_H(t_2+\delta) - \beta_H(t_2)][\beta_H(t_1+\delta) - \beta_H(t_1)] \}$$

$$= \frac{1}{2}\sigma_H^2 \delta^{-2} [(\tau + \delta)^{2H} - 2\tau^{2H} + |\tau - \delta|^{2H}] \qquad (5.6.1\text{-}1)$$

where $\tau = t_2 - t_1$ and $t_2 \geq t_1$. Holding τ fixed and taking the limit as δ approaches zero leaves

$$\phi_{\beta\beta}(\tau) \equiv \sigma_H^2 H(2H-1)|\tau|^{2H-1} \qquad (5.6.1\text{-}2)$$

after applying L'Hopital's Rule twice.

5.6.2 Power Spectral Density of fBm Derivative Process

Because the autocorrelation of real data is an even function, the Fourier transform of any real-valued autocorrelation function may be given in terms of Fourier cosine coefficients instead of the complex exponential coefficients used throughout Chapter 3. Using this method, the PSD of the derivative process is

$$\Phi_{\beta\beta}(f) = \frac{1}{2}\sigma_H^2\,\delta^{-2}\int_{-\infty}^{\infty}[\,(\tau+\delta)^{2H} - 2\tau^{2H} + |\tau-\delta|^{2H}\,]\cos(2\pi f\tau)\,d\tau \quad (5.6.2\text{-}1)$$

which, under the assumption that δf is small, reduces to [26]

$$\Phi_{\beta\beta}(f) = 2\,\sigma_H^2\,\delta^{2H-1}\,\Phi^*_{\beta\beta}(f) \quad (5.6.2\text{-}2)$$

where

$$\Phi^*_{\beta\beta}(f) \equiv K_H(2\pi\delta f)^{1-2H} \quad (5.6.2\text{-}3)$$

$$K_H \equiv \frac{\pi H(2H-1)}{\Gamma(2-2H)}[\cos\pi(H-1)]^{-1} \quad (5.6.2\text{-}4)$$

Thus the fBm derivative process itself may be used to model a system with log-log PSD slope equal to 1-2H. Also, by Equation (4.4-1), fractional Brownian motion models systems with log-log PSD slope equal to -(1+2H).

5.7 Summary of fBm Properties

Fractional Brownian motion is a nonstationary Gaussian process that has stationary, zero mean, Gaussian increments. In contrast with standard Brownian motion, the increments of fBm are not independent of each other. This makes fBm a non-Markov process. The increments of fBm are also self-similar (Equation (5.3-1)). Signals with stationary, self-similar increments are also known as fractals [25].

Some other properties of fBm that are shown in Reference [26] are that it is mean square continuous, almost all of its sample paths are continuous, and it is almost surely not differentiable. Note that these are all properties of Brownian motion, which is the special case of fBm with $H = \frac{1}{2}$.

Reference [26] also shows that fBm is unique in that (1) if a stochastic process has stationary and self-similar increments with parameter H and is mean square continuous, then $0 \le H < 1$, and (2) a Gaussian stochastic process

with self-similar and stationary increments is fractional Brownian motion times a constant plus an offset.

5.8 Simulation of fBm

Just as in the case of Markov stochastic processes, it is useful to be able to simulate fractional Brownian motion using a digital computer. Simulated sample paths are used to test the estimators developed in the Chapter 7. Two methods of simulating fBm are presented below. Simulations were run using both methods, and each produced data with the expected log-log PSD slopes. For the analysis and plots presented in this thesis, the method described in Section 5.8.2 was used.

5.8.1 Discrete Approximation to *fBm* Integral

It is possible to simulate fBm with a brute force discrete approximation to the fBm integral given by Equation (5.2-1). This simulation is implemented as an approximation to the Stratonovich stochastic integral, which is the same as the Ito integral in this case, in the following form:

$$\beta_H(t_k) = \frac{1}{\Gamma(H+.5)}\left\{ \sum_{i=-N}^{-1} \left\{ [t_k - (t_i+\frac{\Delta t}{2})]^{H-1/2} - (-t_i+\frac{\Delta t}{2})^{H-1/2} \right\} [\beta(t_i) - \beta(t_{i-1})] \right.$$

$$\left. + \sum_{i=0}^{k-1} [t_k - (t_i+\frac{\Delta t}{2})]^{H-1/2} [\beta(t_i) - \beta(t_{i-1})] \right\} \qquad (5.8.1\text{-}1)$$

where Δt is the desired time step and $t_i = i\Delta t$. In this formula, N must be very large with respect to k in order to achieve sufficient accuracy. The term $\beta(t_i) - \beta(t_{i-1})$ is a random walk increment which is simulated as $\Delta t^{1/2} w(t_i)$, where $w(t_i)$ is a zero mean, unit variance, Gaussian random variable produced by a random number generator. Multiplying Equation (5.8.1-1) by a

positive constant A will create an fBm sample path with σ_H^2 equal to A^2V_H, where V_H is defined by Equation (5.4-3).

This method is computationally slow because of the necessity to make N large. Experience with the method showed that using N equal to 64k was sufficient to produce good results. However, this requires the generation of 65 times as many Gaussian random variables as there will be points in the resulting fBm sample path.

5.8.2 Correlated Increments Method

An alternate way to simulate fBm is to take advantage of the known correlations between the increments $\beta_H(t_i) - \beta_H(t_{i-1})$. It is possible to generate N independent Gaussian increments and then to transform them so that they have the desired correlations. This is done by creating a matrix which contains the desired correlations between the various increments, i.e. the ij^{th} element of the matrix is given by

$$C_{ij} = \phi_{xx}(n) \qquad (5.8.2\text{-}1)$$

where $\phi_{xx}(n)$ is as defined in Equation (5.5-2), and $n = |i - j|$.

The correlation matrix C may then be factored into the product of a lower triangular symmetric matrix and its transpose using a Cholesky decomposition [40]:

$$C = LL^T \qquad (5.8.2\text{-}2)$$

This is possible because the correlation matrix is positive definite [24]. The positive definiteness of the correlation matrix also guarantees the invertibility of L [40]. It is now possible to define a vector of random variables $\underline{y}$ such that

$$\underline{y} = L^{-1}\underline{x} \qquad (5.8.2\text{-}3)$$

If $\underline{x}$ is zero mean, the variance of $\underline{y}$ is given by

$$E\{\underline{y}\,\underline{y}^T\} = E\{L^{-1}\underline{x}\,\underline{x}^T L^{-T}\}$$

$$= L^{-1} L\, L^T L^{-T}$$

$$= I \qquad (5.8.2\text{-}4)$$

Thus the transformed variables are uncorrelated with unit variance.

The fBm simulation procedure consists of generating the correlation matrix C based upon the desired fBm H and σ_H parameters, factoring C into L and L^T, generating a vector $\underline{y}$ of N independent Gaussian random variables with unit variance, and performing the transformation

$$\underline{x} = L\underline{y} \qquad (5.8.2\text{-}5)$$

to arrive at N fBm increments with the desired correlations. The increments may then be summed to produce the fBm sample path.

This method, which was taken from Reference [24], has the disadvantage of requiring the Cholesky decomposition of the $N \times N$ matrix C. This requires an extremely large number of operations as N becomes large. Even so, this method is still much quicker than the method of Section 5.8.1, and the Cholesky decomposition routine itself is commonly available in software libraries.

Chapter 6

Application of Maximum Likelihood Estimation to System Identification

6.1 Estimation of System Dynamic and Markov Noise Parameters

Maximum likelihood system identification is a method which combines a linear optimal filter with a maximum likelihood estimator to determine unknown dynamic and noise parameters in a given system model. An extended Kalman filter rather than this technique is commonly used to determine system parameters.

Maximum likelihood system identification has been presented under various designations by different authors [15], [29], [31], [37], [39]. It will be called Full Information Maximum Likelihood Optimal Filtering (FIMLOF) in this thesis, a designation commonly used at Draper Laboratory if not in the rest of the community [16], [41]. "Full Information" refers to the fact that stochastic as well as dynamic parameters can be estimated, "Maximum Likelihood" refers to the use of a Fisher estimator to determine the parameters, and "Optimal Filtering" refers to the presence of a Kalman filter in the algorithm.

The basic concept behind FIMLOF is to fix the unknown parameters, propagate the states with a Kalman filter that is based upon those parameters, build a likelihood function with the information obtained from the filter, adjust the parameters to maximize the likelihood function with respect to the unknown parameters, and repeat the process with the new parameter values until the parameter estimates converge.

Alternating between the maximum likelihood parameter estimate iteration and the Kalman filter state propagation is not the same as alternatingly holding some parameters fixed and solving for others in pure maximum likelihood or least squares estimation. The latter procedure is definitely incorrect. The former procedure works very well in practice, because the state initial conditions are among the estimated parameters and the Kalman filter serves to account for noise in the dynamics that maximum likelihood estimation could not handle alone.

6.2 System Model

The system of interest will be assumed to be a linear, time invariant system driven by white noise with white measurement noise. The Ito stochastic differential equation model for the state vector $\underline{x}$ is

$$d\underline{x} = A'\underline{x}\,dt + B'\underline{u}(t)\,dt + L'\,d\underline{\beta}(t) \qquad (6.2\text{-}1)$$

where $\underline{u}(t)$ is a deterministic input vector and $\underline{\beta}(t)$ is a Weiner process. Integration of the equation and discretization yields a discrete time, state space model:

$$\underline{x}(t_{k+1}) = A\underline{x}(t_k) + B\underline{u}(t_k) + L\underline{\xi}(t_k) \qquad (6.2\text{-}2)$$

with measurements

$$\underline{z}(t_{k+1}) = C\underline{x}(t_{k+1}) + \underline{\theta}(t_{k+1}) \qquad (6.2\text{-}3)$$

This model employs the following definitions:

$\underline{x}(t_k) \quad \varepsilon \ \Re^n \quad :$ state vector at time t_k

$\underline{u}(t_k) \quad \varepsilon \ \Re^m \quad :$ deterministic input vector at time t_k

$\underline{\xi}(t_k) \quad \varepsilon \ \Re^p \quad :$ white plant noise vector at time t_k

$\underline{z}(t_k) \quad \varepsilon \ \Re^r \quad :$ measurement vector at time t_k

$\underline{\theta}(t_k) \quad \varepsilon \ \Re^r \quad :$ white measurement noise vector at time t_k

$t_k \quad :$ time index $(k = 0, 1, 2, ...)$

$A \quad \varepsilon \ \Re^{n \times n} \quad :$ state transition matrix

$B \quad \varepsilon \ \Re^{n \times n} \quad :$ system deterministic input matrix

$C \quad \varepsilon \ \Re^{n \times n} \quad :$ system output matrix

$L \quad \varepsilon \ \Re^{n \times n} \quad :$ system plant noise input matrix

The initial state, $\underline{x}(t_0)$, is a Gaussian random variable with mean $E\{\underline{x}(t_0)\} = \bar{\underline{x}}_o$ and covariance $E\{[\underline{x}(t_0) - \bar{\underline{x}}_o][\underline{x}(t_0) - \bar{\underline{x}}_o]^T\} = \Sigma_o$. The plant and measurement noise processes are zero mean, discrete, white noises with covariance $E\{\underline{\xi}(t_j)\underline{\xi}(t_k)\} = \Xi\delta_{jk}$ and $E\{\underline{\theta}(t_j)\underline{\theta}(t_k)\} = \Theta\delta_{jk}$ respectively. This method will also assume that all of the measurement variables are corrupted by the measurement noise and that the input vector $\underline{u}(t_k)$ is known.

6.3 Kalman Filter Equations

The Kalman filter is the most common method of estimating the states of a linear dynamic system. This filter is optimal in the sense that it minimizes the mean square error of the state estimates. The well known equations for the filter are summarized below [14].

$$\Sigma(t_0) = \Sigma_0$$

$$\hat{\underline{x}}(t_0) = \bar{\underline{x}}_o$$

$$\hat{\underline{x}}(t_{k+1} \mid t_k) = A\hat{\underline{x}}(t_k \mid t_k) + B\underline{u}(t_k) \tag{6.3-3}$$

$$\Sigma(t_{k+1} \mid t_k) = A\Sigma(t_k \mid t_k)A^T + L\Xi(t_k)L^T \tag{6.3-4}$$

$$K(t_{k+1}) = \Sigma(t_{k+1} \mid t_k)C^T[C\Sigma(t_{k+1} \mid t_k)C^T + \Theta(t_{k+1})]^{-1} \tag{6.3-5}$$

$$\hat{\underline{x}}(t_{k+1} \mid t_{k+1}) = \hat{\underline{x}}(t_{k+1} \mid t_k) + K(t_{k+1})[\underline{z}(t_{k+1}) - C\hat{\underline{x}}(t_{k+1} \mid t_k)] \tag{6.3-6}$$

$$\Sigma(t_{k+1} \mid t_{k+1}) = [I - K(t_{k+1})C]\,\Sigma(t_{k+1} \mid t_k) \tag{6.3-7}$$

In the formulation above, the notation $\Sigma(t_{k+1} \mid t_k)$ means the covariance of the estimates at time t_{k+1} given all measurements through time t_k. The same notation applies to the estimate of the states, $\hat{\underline{x}}(t_{k+1} \mid t_k)$.

The Kalman filter is characterized by a prediction cycle (Equations 6.3-3 and 6.3-4) and an update cycle (Equations 6.3-6 and 6.3-7). The pre-update residual, or innovation, is defined as

$$r(t_{k+1}) = \underline{z}(t_{k+1}) - \hat{\underline{z}}(t_{k+1} \mid t_k) \tag{6.3-8}$$

where

$$\hat{\underline{z}}(t_{k+1} \mid t_k) = C\hat{\underline{x}}(t_{k+1} \mid t_k) \tag{6.3-9}$$

The mean of this residual is

$$\begin{aligned} E\{\underline{r}(t_{k+1})\} &= E\{\underline{z}(t_{k+1}) - \hat{\underline{z}}(t_{k+1})\} \\ &= E\{C\underline{x}(t_{k+1}) + \underline{\theta}(t_{k+1}) - C\hat{\underline{x}}(t_{k+1} \mid t_k)\} \\ &= C\hat{\underline{x}}(t_{k+1} \mid t_k) - C\hat{\underline{x}}(t_{k+1} \mid t_k) \\ &= 0 \end{aligned} \tag{6.3-10}$$

The covariance $S(t_{k+1})$ of the pre-update residual is

$$S(t_{k+1}) \equiv E\{\underline{r}(t_{k+1})\,\underline{r}(t_{k+1})^T\}$$

$$= E\{[C\underline{x}(t_{k+1}) + \underline{\theta}(t_{k+1}) - C\hat{\underline{x}}(t_{k+1} \mid t_k)][C\underline{x}(t_{k+1}) + \underline{\theta}(t_{k+1}) - C\hat{\underline{x}}(t_{k+1} \mid t_k)]^T\}$$

$$= C\Sigma(t_{k+1} \mid t_k)C^T + \Theta(t_{k+1}) \qquad (6.3\text{-}11)$$

In Section 6.4, it will be useful to know the conditional probability density of the measurement $\underline{z}(t_k)$ at time t_k given all measurements $\underline{z}^{k-1} = [\underline{z}(t_{k-1}), \underline{z}(t_{k-2}), \ldots, \underline{z}(t_0)]$ up to t_{k-1}, where $\underline{z}(t_0)$ contains the initial condition information. Using Equations (6.3-10) and (6.3-11) along with the assumption that all of the noise processes are Gaussian, this density may be written as

$$p(\underline{z}(t_k) \mid \underline{z}^{k-1}) = (2\pi)^{-r/2}\det[S(t_k)]^{-1/2}\, e^{-[\underline{r}(t_k)^T S(t_k)^{-1}\underline{r}(t_k)]/2} \qquad (6.3\text{-}12)$$

where det(S) is the determinant of the matrix S.

6.4 The Likelihood Function

As stated in Chapter 2, the goal of maximum likelihood estimation is to maximize the joint probability density of the measurements with respect to the unknown parameters. In this way, the parameter values are chosen that maximize the probability that the measurements that did occur would have occurred.

The notation $p(\underline{z}^N;\underline{\alpha})$ indicates the joint probability density function of all of the measurements through time t_N. The vector $\underline{\alpha}$ is composed of the unknown parameters. The expression $p(\underline{z}^N;\underline{\alpha})$ is actually a family of densities that are indexed by the different values of the parameters [37]. The probability density may be expanded using conditional densities

$$p(\underline{z}^N;\underline{\alpha}) = p(\underline{z}(t_N) \mid \underline{z}^{N-1};\underline{\alpha}) \ldots p(\underline{z}(t_1) \mid \underline{z}(t_0);\underline{\alpha})\, p(\underline{z}(t_0);\underline{\alpha}) \qquad (6.4\text{-}1)$$

Since $p(\underline{z}(t_N) \mid \underline{z}^{N-1};\underline{\alpha})$ as given by Equation (6.3-12) is a function of $\underline{r}(t_k)$ and $S(t_k)$ only, the likelihood function may be computed using quantities calculated by the Kalman filter.

It will simplify the computations if $\ln[p(\underline{z}^N;\underline{\alpha})]$ is maximized in place of the density itself. This eliminates the exponential and converts the products

into sums. This simplification is made possible by the fact that the natural logarithm increases monotonically. By Equation (6.4-1),

$$\ln[p(\underline{z}^N;\underline{\alpha})] = \sum_{k=0}^{N} \ln[p(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha})] \qquad (6.4\text{-}2)$$

where $p(\underline{z}(0) \mid \underline{z}^{-1};\underline{\alpha}) \equiv p(\underline{z}(0);\underline{\alpha})$.

Using Equation (6.3-12)

$$\ln[p(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha})] = -\frac{r}{2}\ln(2\pi) - \frac{1}{2}\ln\{\det[S(t_k;\underline{\alpha})]\} - \frac{1}{2}\underline{r}(t_k;\underline{\alpha})^T S(t_k;\underline{\alpha})^{-1}\underline{r}(t_k;\underline{\alpha}) \qquad (6.4\text{-}3)$$

Because the leading term is a constant, it will not affect the parameters that maximize the function and it can be ignored. Finally, changing the sign of the remaining terms produces the negative log-likelihood function without the constant term:

$$\zeta(\underline{z}^N;\underline{\alpha}) \equiv \sum_{k=0}^{N} \zeta(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha}) \qquad (6.4\text{-}4)$$

where

$$\zeta(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha}) \equiv \frac{1}{2}\ln\{\det[S(t_k;\underline{\alpha})]\} + \frac{1}{2}\underline{r}(t_k;\underline{\alpha})^T S(t_k;\underline{\alpha})^{-1}\underline{r}(t_k;\underline{\alpha}) \qquad (6.4\text{-}5)$$

This negative log-likelihood function may be assembled recursively as the Kalman filter processes the measurements. The maximum likelihood estimate $\hat{\underline{\alpha}}$ of the parameters $\underline{\alpha}$ will be the parameter values which minimize Equation (6.4-4).

6.5 Minimizing the Negative Log-Likelihood Function

There are several numerical methods available to minimize the negative log-likelihood function. Newton-Raphson iteration is chosen because it offers relatively fast convergence and because it is possible to analytically determine the gradient and an approximation to the Hessian of $\zeta(\underline{z}^N;\underline{\alpha})$.

Let $\underline{\hat{\alpha}} = [\hat{\alpha}_1, ..., \hat{\alpha}_q]^T$ be the maximum likelihood estimates of the parameters. For these values the negative log-likelihood function is minimized

$$\zeta(\underline{z}^N;\underline{\hat{\alpha}}) \quad = \quad \text{minimum} \tag{6.5-1}$$

This means that the gradient of the function evaluated using these values for the parameters is zero

$$\frac{\partial \zeta(\underline{z}^N;\underline{\alpha})}{\partial \alpha_i}\Bigg|_{\underline{\alpha}=\underline{\hat{\alpha}}} \quad = \quad 0 \tag{6.5-2}$$

If $\underline{\alpha}_o = [\alpha_{1o},...,\alpha_{qo}]^T$ are the first guesses at the values of the parameters and if Equation (6.5-2) is expanded in a Taylor series about these guesses, the result is

$$0 \quad = \quad \frac{\partial \zeta(\underline{z}^N;\alpha)}{\partial \alpha_i}\Bigg|_{\underline{\alpha}=\underline{\hat{\alpha}}}$$

$$= \quad \frac{\partial \zeta(\underline{z}^N;\underline{\alpha})}{\partial \alpha_i}\Bigg|_{\underline{\alpha}=\underline{\alpha}_o} + \sum_{j=1}^{q} \frac{\partial^2 \zeta(\underline{z}^N;\underline{\alpha})}{\partial \alpha_i \partial \alpha_j}\Bigg|_{\underline{\alpha}=\underline{\alpha}_o} \Delta\alpha_j, \quad i = 1, ..., q \tag{6.5-3}$$

where $\Delta\alpha_j \quad = \quad \hat{\alpha}_j - \alpha_{jo}$.

As in Section 2.6, this result leads to the following set of linear equations for the adjustments $\Delta\alpha_j$ to the first guesses α_{jo} towards the maximum likelihood estimates $\hat{\alpha}_j$:

$$\sum_{j=1}^{q} A_{ij}\,\Delta\alpha_j \quad = \quad B_i, \qquad i \quad = \quad 1, ..., q \tag{6.5-4}$$

where by Equations (6.4-4) and (6.5-3)

$$B_i \quad = \quad -\sum_{k=0}^{N} \frac{\partial \zeta(\underline{z}(k) \mid \underline{z}^{k-1};\underline{\alpha})}{\partial \alpha_i}\Bigg|_{\underline{\alpha}=\underline{\alpha}_o}, \quad i = 1, ..., q \tag{6.5-5}$$

$$A_{ij} = \sum_{k=0}^{N} \frac{\partial^2 \zeta(\underline{z}(k) \mid \underline{z}^{k-1};\underline{\alpha})}{\partial\alpha_i \partial\alpha_j} \Bigg|_{\underline{\alpha}=\underline{\alpha}_0} , \; i,j = 1, \ldots, q \qquad (6.5\text{-}6)$$

In the equations above, B and A are the negative gradient and the Hessian of the negative log-likelihood function respectively. The maximum likelihood estimate of the parameters is found by solving Equation (6.5-4) iteratively until $\Delta\underline{\alpha}$ approaches zero.

6.6 Derivatives of the Likelihood Function

In order to use the preceding minimization method the first partial derivatives of the negative log-likelihood function with respect to the unknown parameters are calculated analytically. These derivatives are the gradient and they are used to form an approximation to the Hessian. The gradient follows directly from Equations (6.4-4) and (6.4-5). It is

$$\frac{\partial \zeta(\underline{z}^N;\underline{\alpha})}{\partial\alpha_i} = \sum_{k=0}^{N} \frac{\partial \zeta(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha})}{\partial\alpha_i} , \quad i = 1, \ldots, q \qquad (6.6\text{-}1)$$

where

$$\begin{aligned} \frac{\partial \zeta(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha})}{\partial\alpha_i} &= \underline{r}(t_k;\underline{\alpha})^T S(t_k;\underline{\alpha})^{-1} \frac{\partial \underline{r}(t_k;\underline{\alpha})}{\partial\alpha_i} \\ &\quad - \frac{1}{2} \underline{r}(t_k;\underline{\alpha})^T S(t_k;\underline{\alpha})^{-1} \frac{\partial S(t_k;\underline{\alpha})}{\partial\alpha_i} S(t_k;\underline{\alpha})^{-1} \underline{r}(t_k;\underline{\alpha}) \\ &\quad + \frac{1}{2} \operatorname{tr}\left[S(t_k;\underline{\alpha})^{-1} \frac{\partial S(t_k;\underline{\alpha})}{\partial\alpha_i} \right] \end{aligned} \qquad (6.6\text{-}2)$$

Appendix B contains the derivation of the last term which is the derivative of the natural logarithm of the determinant.

If the definition of $\underline{r}$ and the formula for S, Equation (6.3-11), are substituted into Equation (6.6-2), the gradient may be calculated analytically without much difficulty.

In order to avoid calculating the second partial derivatives of the negative log-likelihood function, the Fisher information matrix is used as an approximation to the Hessian. Recall from Section 2.3 that the Fisher information matrix is equal to the expected value of the Hessian. Assuming the Hessian is approximately equal to its expected value, Equation (2.3-2) may be used to express Equation (6.5-6) as

$$A_{ij} \cong \sum_{k=0}^{N} E\left\{ \frac{\partial \zeta(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha})}{\partial \alpha_i} \frac{\partial \zeta(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha})}{\partial \alpha_j} \right\} \qquad (6.6\text{-}3)$$

This Fisher information matrix approximation is valid when the deterministic input is very large with respect to the stochastic input to the system or when the observation interval $[0, t_N]$ is much greater than the correlation times of $\frac{\partial \underline{r}(t_k;\underline{\alpha})}{\partial \alpha_i}$ and $\frac{\partial \underline{r}(t_k;\underline{\alpha})}{\partial \alpha_j}$ [37].

Equation (6.6-3) may be manipulated to arrive at the more useful form given by Equation (6.6-4). The derivation of this equation in the special case of a scalar measurement is presented in Section 6.8. The derivation of the general form is contained in Appendix C.

$$A_{ij} \cong \sum_{k=0}^{N} \mathrm{tr}\left[\frac{\partial \underline{r}(t_k;\underline{\alpha})}{\partial \alpha_i} \frac{\partial \underline{r}(t_k;\underline{\alpha})^T}{\partial \alpha_j} S(t_k;\underline{\alpha})^{-1} + \frac{1}{2} \frac{\partial S(t_k;\underline{\alpha})}{\partial \alpha_i} S(t_k;\underline{\alpha})^{-1} \frac{\partial S(t_k;\underline{\alpha})}{\partial \alpha_j} S(t_k;\underline{\alpha})^{-1} \right] \qquad (6.6\text{-}4)$$

The Fisher information approximation allows the use of a Newton-Raphson iteration which does not require the difficult task of calculating the second partial derivatives of the negative log-likelihood function. This method is known as Gauss-Newton iteration.

There is evidence that the Fisher information approximation is better than using the exact Hessian. Namely, the second partial derivatives were coded for the model in Section 6.8. The maximum likelihood iteration did not converge with the exact Hessian, but it did converge with the Fisher information approximation to the Hessian. The computer code for the first and second partial derivatives of the negative log-likelihood function was checked by the difference method (Appendix D.4), so this result is probably

not due to a coding error. Statistical variations apparently make the exact Hessian less tractable than the expected value of the Hessian, at least when the first guess for the parameters is not close to the true values. For example, the Hessian may have some negative eigenvalues, while the Fisher information approximation is guaranteed to be positive definite. See also References [15], [41]. In this thesis, all minimizations were performed using a pure Gauss-Newton iteration.

6.7 Implementation

The FIMLOF software algorithm consists of implementing the Kalman filter described by Equations (6.3-3) through (6.3-7). Using a first guess for the parameters, this filter calculates the residual and error covariance at each measurement. These values, along with their partial derivatives are then incorporated into the negative log-likelihood function, its gradient, and the approximation to its Hessian. Once all the measurements have been processed, a single Gauss-Newton iteration is performed to update the parameter estimates. The process is repeated with a new Kalman filter which is based upon the improved parameter estimates, and iterations continue until the parameters converge satisfactorily.

FORTRAN software was written to implement the FIMLOF algorithm. Much of the code is for systems of general dimensions, but some subroutines were specialized to scalar rather than vector observables and to the specific two state model of Section 6.8. The software uses the square roots of the diagonal elements of the Fisher information matrix (calculated during the maximum likelihood iteration) as a measure of the uncertainty of the parameter estimates, by the Cramer-Rao lower bound discussed in Section 2.4.

It can be shown that for a time invariant, linear Gaussian model with stationary noise processes, the FIMLOF estimate will display the properties which make classical maximum likelihood estimates attractive [37]. That is, the FIMLOF estimate is asymptotically consistent, unbiased, normally distributed, and efficient. As stated in Chapter 2, this means that the Cramer-

Rao lower bound is attained in the limit of a large number of observations. Thus the accuracy of the parameter estimates can be readily evaluated.

6.8 Application of FIMLOF to an Approximation to fBm

6.8.1 Example of Experimental Data with a -1 log-log PSD Slope

As a nonMarkov process, fractional Brownian motion is not suitable for estimation using a Kalman filter, since the derivation of the Kalman filter assumes Markov plant noise [18]. This means that FIMLOF cannot be applied to estimate fBm parameters. However, it is possible to produce a system which has a power spectral density with a -1 slope over a finite frequency range by summing several Markov noise processes [21]. An example of this procedure using experimental data is now presented.

Figure 6.8-1 shows the output of an accelerometer which is under a constant -1 g acceleration. The plot was produced by sampling the output at 1 Hz and then averaging every 10 data points. The data covers a time span of 5.7 hours. Figure 6.8-2 is a PSD of the 1 Hz data using 32,768 (2^{15}) points. Below 0.05 Hz, this plot has the -1 log-log slope which is characteristic of fBm. Above this frequency, the plot transitions through a zero slope to a +2 log-log slope. A slope of zero is indicative of the presence of white noise and a +2 slope is an attribute of quantization noise. Figure 6.8-3 is a PSD of the 10 second averaged data. It shows that the averaging removes some of the quantization noise while adding some energy to the lower frequencies due to aliasing. The estimator will use this averaged data in order to decrease the number of measurements while still covering a long time span.

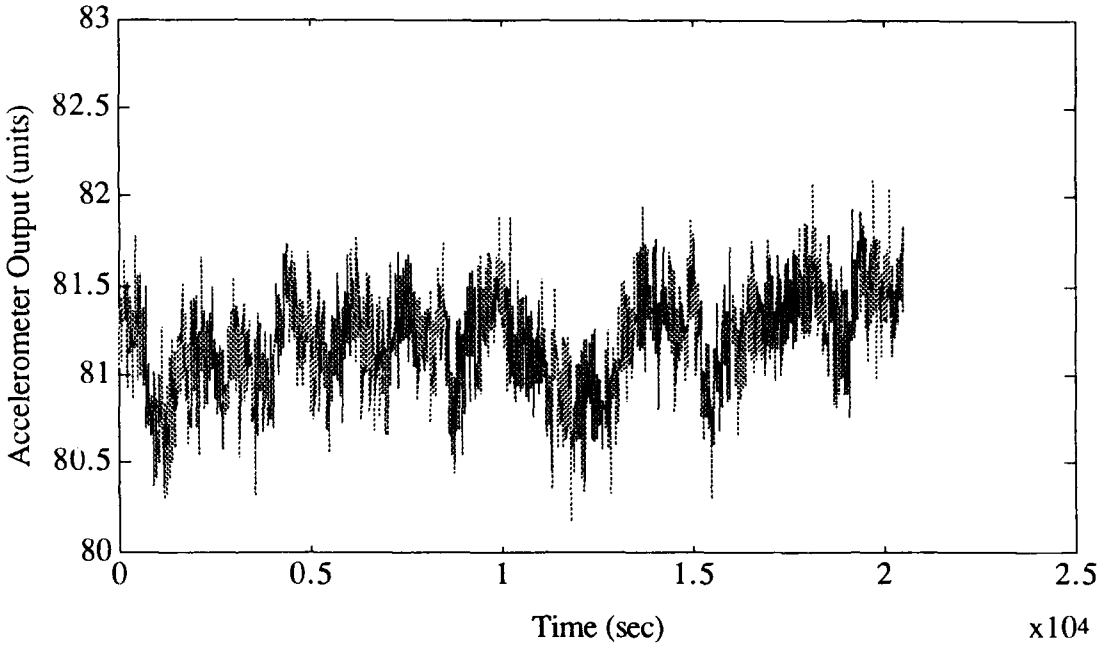


Figure 6.8-1 Accelerometer Output Averaging 10 sec Intervals

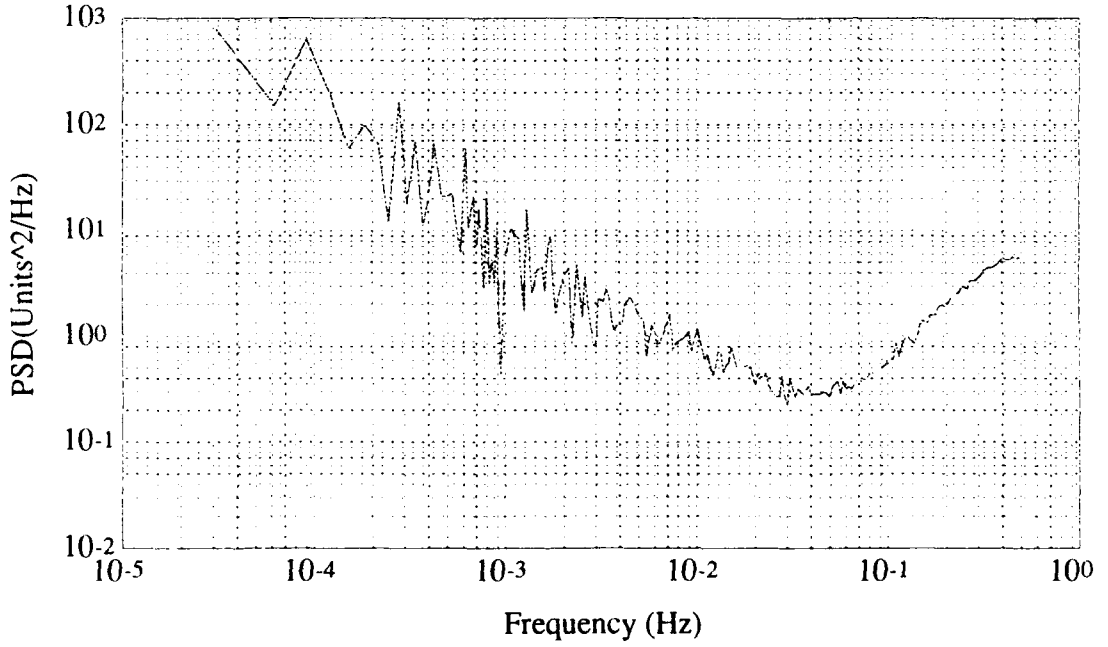


Figure 6.8-2 PSD of Accelerometer Output Using 32,768 Points

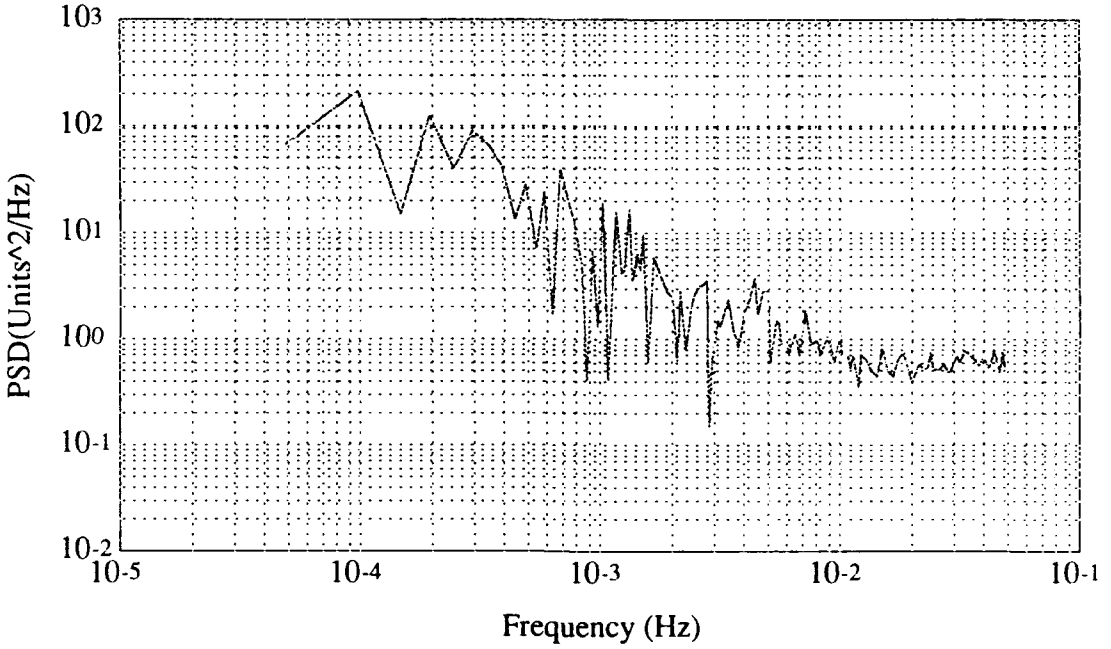


Figure 6.8-3 PSD of Accelerometer Output Using 2048 Points of 10 sec Averaged Data

6.8.2 Second Order Scalar Observable FIMLOF Model

The fact that this problem involves a scalar measurement simplifies the calculations and makes it a good example to more clearly demonstrate the FIMLOF process. The fractional Brownian motion contained in the signal will be approximated by the combination of a trend, random walk, and exponentially correlated noise which are all Markov noise processes. The system model will also include white measurement noise.

Let (x_1, x_2) be states with x_1 equal to a trend plus random walk and x_2 equal to an exponentially correlated noise. Let the deterministic input function $u(t_k)$ be a unit step, and assume that $\Delta t = t_k - t_{k-1}$ is the constant time spacing of the measurements in seconds. Also make the following definitions:

a : trend parameter
b : random walk standard deviation
c_1 : inverse of the exponentially correlated noise time constant
c_2 : exponentially correlated noise scaling parameter
v : white noise standard deviation

Then following the notation of Section 6.2, this model is written as

$$\underline{x}(t_{k+1}) = \begin{bmatrix} 1 & 0 \\ 0 & e^{-c_1\Delta t} \end{bmatrix} \underline{x}(t_k) + \begin{bmatrix} a\Delta t \\ 0 \end{bmatrix} u(t_k) + \begin{bmatrix} b & 0 \\ 0 & c_1c_2 \end{bmatrix} \underline{\xi}(t_k) \quad (6.8.2\text{-}1)$$

$$\underline{z}(t_{k+1}) = [1\ 1]\, \underline{x}(t_{k+1}) + \underline{\theta}(t_{k+1}) \quad (6.8.2\text{-}2)$$

$$\Xi(t_k) = \Delta t \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (6.8.2\text{-}3)$$

$$\Theta(t_k) = [v^2] \quad (6.8.2\text{-}4)$$

The state equation (6.8.2-1) for x_2 is the discrete form of Equation (4.8-1).

Based upon the above formulation, the parameters to be estimated are a, b, c_1, c_2, and v. Four more parameters are needed because the Kalman filter requires initial conditions on the state estimates and their covariance.

$$\bar{\underline{x}}_0 = \begin{bmatrix} x_{10} \\ x_{20} \end{bmatrix} \quad (6.8.2\text{-}5)$$

$$\Sigma_0 = \begin{bmatrix} s_1^2 & 0 \\ 0 & s_2^2 \end{bmatrix} \quad (6.8.2\text{-}6)$$

The initial covariance matrix is assumed diagonal because the noise states are independent in the model. The initial conditions are estimated along with a, b, c_1, c_2, and v, but s_1 and s_2 are assumed to have known values. The derivation of the estimator will include the formulas involving s_1 and s_2 for completeness.

The Kalman filter equations for this system are well defined and follow directly from those listed in Section 6.3. In order to perform the minimization of the negative log-likelihood function using the gradient and approximation to the Hessian which are given by Equations (6.6-2) and (6.6-4) the partial derivatives of $r(t_{k+1};\underline{\alpha})$, which is a scalar, and $S(t_{k+1};\underline{\alpha})$ are required.

Expressions for these partial derivatives are formed using the Kalman filter equations which are specific to the system.

Recall from Equation (6.3-11) that in the scalar measurement case the error covariance of the residuals is

$$\begin{aligned} S(t_{k+1}) &= E\{ r(t_{k+1})^2 \} \\ &= \Sigma_{11}(t_{k+1} \mid t_k) + \Sigma_{12}(t_{k+1} \mid t_k) \\ &\quad + \Sigma_{21}(t_{k+1} \mid t_k) + \Sigma_{22}(t_{k+1} \mid t_k) + v^2 \end{aligned} \tag{6.8.2-7}$$

where the explicit dependence of $\underline{r}$ and S on $\underline{\alpha}$ has been dropped for clarity. Then the Kalman filter gain matrix is

$$K(t_{k+1}) = \Sigma(t_{k+1} \mid t_k)\, C^T [\, C\, \Sigma(t_{k+1} \mid t_k)\, C^T + v^2 \,]^{-1}$$

$$= \frac{1}{S(t_{k+1})} \begin{bmatrix} \Sigma_{11} + \Sigma_{12} \\ \Sigma_{21} + \Sigma_{22} \end{bmatrix} \tag{6.8.2-8}$$

The covariance matrix of the updated state is

$$\Sigma(t_{k+1} \mid t_{k+1}) = [\, I - K(t_{k+1})\, C \,]\, \Sigma(t_{k+1} \mid t_k) \tag{6.8.2-9}$$

$$= \Sigma(t_{k+1} \mid t_k) - \Sigma(t_{k+1} \mid t_k)\, C^T [\, C\, \Sigma(t_{k+1} \mid t_k)\, C^T + v^2 \,]^{-1}\, C\, \Sigma(t_{k+1} \mid t_k)$$

and Kalman filter update Equation (6.3-6) is

$$\hat{\underline{x}}(t_{k+1} \mid t_{k+1}) = \hat{\underline{x}}(t_{k+1} \mid t_k) + K(t_{k+1}) r(t_{k+1}) \tag{6.8.2-10}$$

At the initial time t_o, the only non-zero partial derivatives are by Equations (6.8.2-5) and (6.8.2-6)

$$\frac{\partial \hat{x}_1(t_0)}{\partial x_{10}} = 1, \qquad \frac{\partial \hat{x}_2(t_0)}{\partial x_{20}} = 1 \tag{6.8.2-11}$$

$$\frac{\partial \Sigma_{11}(t_0)}{\partial s_1} = 2\, s_1, \qquad \frac{\partial \Sigma_{22}(t_0)}{\partial s_2} = 2\, s_2 \tag{6.8.2-12}$$

The only non-zero partial derivative of the state transition matrix is

$$\frac{\partial A}{\partial c_1} = \begin{bmatrix} 0 & 0 \\ 0 & -\Delta t e^{-c_1 \Delta t} \end{bmatrix} \tag{6.8.2-13}$$

The partial derivatives of the state propagated from time t_k to t_{k+1} with respect to parameters, initial conditions, or initial state standard deviations α_i are by Equation (6.3-3)

$$\frac{\partial \hat{\underline{x}}(t_{k+1} \mid t_k)}{\partial \alpha_i} = \frac{\partial A}{\partial \alpha_i} \hat{\underline{x}}(t_k \mid t_k) + A \frac{\partial \hat{\underline{x}}(t_k \mid t_k)}{\partial \alpha_i} + \frac{\partial B}{\partial \alpha_i} \tag{6.8.2-14}$$

where the only non-zero partial derivative of B is

$$\frac{\partial B}{\partial a} = \begin{bmatrix} \Delta t \\ 0 \end{bmatrix} \tag{6.8.2-15}$$

The partial derivatives of the state covariance propagated from time t_k to t_{k+1} with respect to parameters, initial conditions, or initial state standard deviations α_i are by Equation(6.3-4)

$$\begin{aligned} \frac{\partial \Sigma(t_{k+1} \mid t_k)}{\partial \alpha_i} = & \frac{\partial A}{\partial \alpha_i} \Sigma(t_k \mid t_k) A^T + A \Sigma(t_k \mid t_k) \frac{\partial A^T}{\partial \alpha_i} \\ & + A \frac{\partial \Sigma(t_k \mid t_k)}{\partial \alpha_i} A^T + \frac{\partial Q}{\partial \alpha_i} \end{aligned} \tag{6.8.2-16}$$

where

$$Q \equiv L \Xi L^T = L L^T \Delta t \tag{6.8.2-17}$$

so that

$$\frac{\partial Q}{\partial \alpha_i} = \frac{\partial L}{\partial \alpha_i} L^T \Delta t + L \frac{\partial L^T}{\partial \alpha_i} \Delta t \tag{6.8.2-18}$$

with the only non-zero partial derivatives of L being

$$\frac{\partial L}{\partial b} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \frac{\partial L}{\partial c_1} = \begin{bmatrix} 0 & 0 \\ 0 & c_2 \end{bmatrix}, \quad \frac{\partial L}{\partial c_2} = \begin{bmatrix} 0 & 0 \\ 0 & c_1 \end{bmatrix} \tag{6.8.2-19}$$

Also the partial derivative of the measurement noise covariance is

$$\frac{\partial v^2}{\partial \alpha_i} = \begin{cases} 0 & \text{if } \alpha_i \neq v \\ 2v & \text{if } \alpha_i = v \end{cases} \tag{6.8.2-20}$$

At this point, recalling the definitions of the residual and its covariance, their partial derivatives may be written as

$$\frac{\partial\, r(t_{k+1})}{\partial \alpha_i} = -C \frac{\partial \hat{\underline{x}}(t_{k+1} \mid t_k)}{\partial \alpha_i} \tag{6.8.2-21}$$

$$\frac{\partial\, S(t_{k+1})}{\partial \alpha_i} = \frac{\partial \Sigma_{11}(t_{k+1} \mid t_k)}{\partial \alpha_i} + \frac{2\,\partial \Sigma_{12}(t_{k+1} \mid t_k)}{\partial \alpha_i} + \frac{\partial \Sigma_{22}(t_{k+1} \mid t_k)}{\partial \alpha_i} + 2\,v \frac{\partial v}{\partial \alpha_i} \tag{6.8.2-22}$$

The partial derivatives of the Kalman filter gain matrix are by Equation (6.8.2-8)

$$\frac{\partial K(t_{k+1})}{\partial \alpha_i} = \frac{1}{S(t_{k+1})} \begin{bmatrix} \frac{\partial \Sigma_{11}(t_{k+1} \mid t_k)}{\partial \alpha_i} + \frac{\partial \Sigma_{12}(t_{k+1} \mid t_k)}{\partial \alpha_i} \\ \frac{\partial \Sigma_{21}(t_{k+1} \mid t_k)}{\partial \alpha_i} + \frac{\partial \Sigma_{22}(t_{k+1} \mid t_k)}{\partial \alpha_i} \end{bmatrix} - \frac{1}{S(t_{k+1})^2} \frac{\partial S(t_{k+1})}{\partial \alpha_i} \begin{bmatrix} \Sigma_{11}(t_{k+1} \mid t_k) + \Sigma_{12}(t_{k+1} \mid t_k) \\ \Sigma_{21}(t_{k+1} \mid t_k) + \Sigma_{22}(t_{k+1} \mid t_k) \end{bmatrix} \tag{6.8-23}$$

The partial derivatives of the updated state at time t_{k+1} given the measurements up to time t_{k+1} are by Equation (6.8.2-10)

$$\frac{\partial \hat{\underline{x}}(t_{k+1} \mid t_{k+1})}{\partial \alpha_i} = \frac{\partial \hat{\underline{x}}(t_{k+1} \mid t_k)}{\partial \alpha_i} + K(t_{k+1}) \frac{\partial r(t_{k+1})}{\partial \alpha_i} + \frac{\partial K(t_{k+1})}{\partial \alpha_i} r(t_{k+1}) \tag{6.8-24}$$

The partial derivatives of the updated state covariance at time t_{k+1} given the measurements up to time t_{k+1} are by Equation (6.8.2-9)

$$\frac{\partial \Sigma(t_{k+1} \mid t_{k+1})}{\partial \alpha_i} = \frac{\partial \Sigma(t_{k+1} \mid t_k)}{\partial \alpha_i} - K(t_{k+1})\, C \frac{\partial \Sigma(t_{k+1} \mid t_k)}{\partial \alpha_i} - \frac{\partial K(t_{k+1})}{\partial \alpha_i} C\, \Sigma(t_{k+1} \mid t_k) \qquad (6.8.2\text{-}25)$$

In this simplified case the components of the gradient of the negative log-likelihood function (6.6-2) may now be written as

$$\frac{\partial \zeta(z(t_{k+1}) \mid z^k;\alpha)}{\partial \alpha_i} = \frac{1}{2S(t_{k+1})} \frac{\partial S(t_{k+1})}{\partial \alpha_i} + \frac{r(t_{k+1})}{S(t_{k+1})} \frac{\partial r(t_{k+1})}{\partial \alpha_i} - \frac{r(t_{k+1})^2}{2\, S(t_{k+1})^2} \frac{\partial S(t_{k+1})}{\partial \alpha_i} \qquad (6.8.2\text{-}26)$$

and the Hessian (6.5-6) of the negative log-likelihood function is

$$A_{ij} = \sum_{k=0}^{N} A_{ij}(t_k) \qquad (6.8.2\text{-}27)$$

with

$$A_{ij}(t_k) = \frac{\partial^2 \zeta(z(t_k) \mid z^{k-1};\alpha)}{\partial \alpha_i \partial \alpha_j} \qquad (6.8.2\text{-}28)$$

Inserting Equations (6.8.2-26) into Equation (6.6-3) yields the Fisher information approximation to the Hessian

$$\begin{aligned} A_{ij}(t_k) \;&\cong\; E\{A_{ij}(t_k)\} \\ &= E\Bigg\{ \frac{1}{4\, S(t_k)^2} \frac{\partial S(t_k)}{\partial \alpha_i} \frac{\partial S(t_k)}{\partial \alpha_j} + \frac{r(t_k)}{2\, S(t_k)^2} \Bigg[\frac{\partial S(t_k)}{\partial \alpha_i} \frac{\partial r(t_k)}{\partial \alpha_j} + \frac{\partial S(t_k)}{\partial \alpha_j} \frac{\partial r(t_k)}{\partial \alpha_i} \Bigg] + \frac{r(t_k)^2}{S(t_k)^2} \frac{\partial r(t_k)}{\partial \alpha_i} \frac{\partial r(t_k)}{\partial \alpha_j} \end{aligned}$$

$$+ \frac{r(t_k)^4}{4\,S(t_k)^4} \frac{\partial S(t_k)}{\partial \alpha_i} \frac{\partial S(t_k)}{\partial \alpha_j}$$
$$- \frac{r(t_k)^2}{2\,S(t_k)^3} \frac{\partial S(t_k)}{\partial \alpha_i} \frac{\partial S(t_k)}{\partial \alpha_j}$$
$$- \frac{r(t_k)^3}{2\,S(t_k)^3} \left[\frac{\partial S(t_k)}{\partial \alpha_i} \frac{\partial r(t_k)}{\partial \alpha_j} + \frac{\partial S(t_k)}{\partial \alpha_j} \frac{\partial r(t_k)}{\partial \alpha_i} \right] \Bigg\} \qquad (6.8.2\text{-}29)$$

If the following facts are applied to the above equation

$r(t_k)$: zero mean Gaussian distribution with
covariance = $S(t_{k+1})$
third moment = 0
fourth moment = $3\,S(t_{k+1})^2$

$\frac{\partial r(t_k)}{\partial \alpha_i}$: non-random function relative to $y_k \mid y_{k-1},\ldots,y_0$

$S(t_k)$: non-random function relative to $y_{k+1} \mid y_k,\ldots,y_1$

it reduces to

$$E\{A_{ij}(t_k)\} = \frac{1}{2\,S(t_k)^2} \frac{\partial S(t_k)}{\partial \alpha_i} \frac{\partial S(t_k)}{\partial \alpha_j} + \frac{1}{S(t_k)} \frac{\partial r(t_k)}{\partial \alpha_i} \frac{\partial r(t_k)}{\partial \alpha_i} \qquad (6.8.2\text{-}30)$$

so that

$$A_{ij} \cong \sum_{k=0}^{N} \left\{ \frac{1}{2\,S(t_k)^2} \frac{\partial S(t_k)}{\partial \alpha_i} \frac{\partial S(t_k)}{\partial \alpha_j} + \frac{1}{S(t_k)} \frac{\partial r(t_k)}{\partial \alpha_i} \frac{\partial r(t_k)}{\partial \alpha_i} \right\} \qquad (6.8.2\text{-}31)$$

As expected, Equation (6.8.2-31) is the scalar version of Equation (6.6-4).

6.8.3 Application of FIMLOF to Experimental Data

The FIMLOF estimation software described in Section 6.7 for the second order scalar observable model of Section 6.8.2 was applied to the 10 point averaged data of Figure 6.8.1. Two fits were performed using different time spans of data. One fit covered 30,000 seconds and the other 48,500 seconds. The results of these fits are contained in Table 6.8-1.

Table 6.8-1 Results of FIMLOF Fit to Accelerometer Data

	30,000 sec of Data	48,500 sec of Data
Initial Conditions: x_1 (units) x_2 (units)	 0.00 ± 10.0 -0.158 ± 1.0	 0.10 ± 10.0 0.002 ± 1.0
Trend: a (units/sec)	$2.09 \times 10^{-5} \pm 2.8 \times 10^{-5}$	$7.94 \times 10^{-7} \pm 2.7 \times 10^{-5}$
Random Walk Standard Deviation: b (units)	$5.10 \times 10^{-3} \pm 1.2 \times 10^{-3}$	$5.89 \times 10^{-3} \pm 1.0 \times 10^{-3}$
Exp. Corr. Noise: Inv. Time Const.: c_1 (1/sec) Scaling Param.: c_2 (units)	 $5.74 \times 10^{-3} \pm 1.5 \times 10^{-3}$ 2.70 ± 0.86	 $6.59 \times 10^{-3} \pm 1.4 \times 10^{-3}$ 2.76 ± 0.54
White Noise Standard Deviation: v (units)	0.174 ± 0.003	0.173 ± 0.002
RMS Pre-update Residual: (units)	0.20	0.20

These numerical results present some information about the quality of the fit. It is important to note that the addition of more measurements decreased the uncertainty of the estimates as measured by the Cramer-Rao lower bound. This lower bound should improve by a factor of the square root of the ratio of the first to the second number of measurements. The extremely low estimate of the trend and the associated high uncertainty indicates that there is no significant trend over the time span of the data.

Because the goal was to approximate a PSD with a -1 log-log slope, a better test of the quality of the fit occurs in the frequency domain. Simulated sample paths were produced with the techniques described in Section 4.8. These simulations of the system model (6.8.2-1) used the parameters given by the 48,500 second fit. Figure 6.8-4 is an example of one such path and Figure 6.8-5 shows a PSD generated from the data using 2,048 points. The PSD is nearly identical to Figure 6.8-3 which shows the frequency content of the averaged data. This is reasonable because of the measurement averaging that was used to reduce the number of data points.

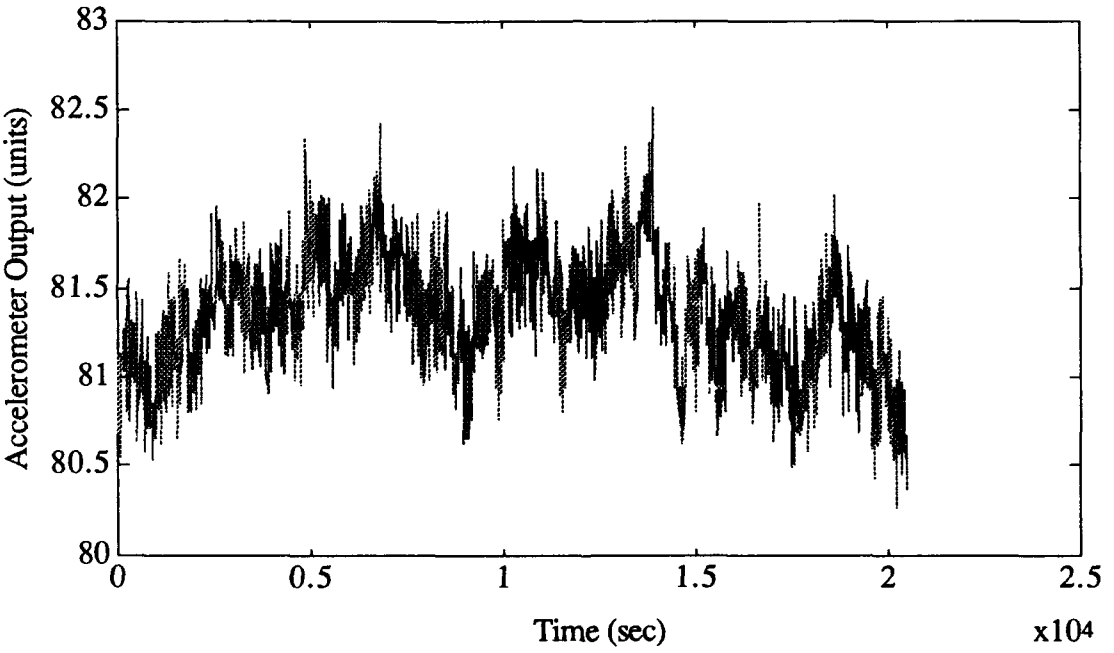


Figure 6.8-4 Simulated Sample Path

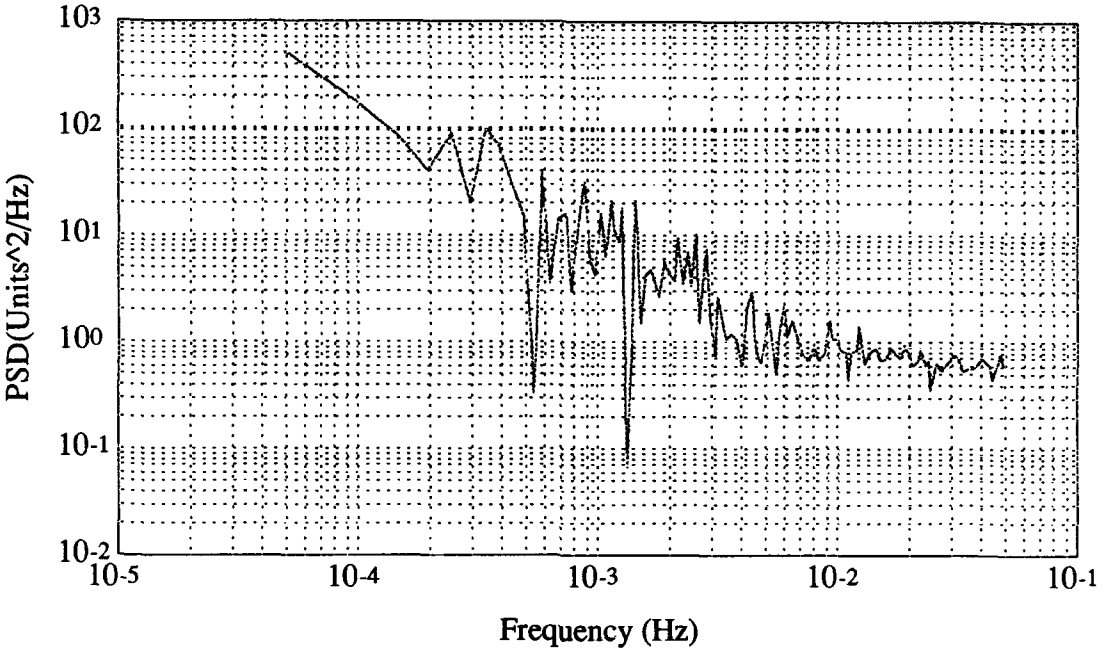


Figure 6.8-5 PSD of Simulated Sample Path

This example shows that FIMLOF may be used to identify a Markov model which will approximate the power spectrum of a fractional Brownian motion process over a given frequency range. However, a PSD produced using a longer time span of data would show flattening in the lower frequencies. Increasing the valid range beyond the 3 decades presented in this example requires that more states be added to the model. Reference [21] provides an analysis of the number of poles required to approximate a -1 slope over a given number of decades.

6.9 Results of Fits to Computer Generated Sample Paths

In order to further demonstrate the accuracy of FIMLOF, a sample path consisting of the sum of a linear trend, random walk, exponentially correlated noise, and white noise was generated using the method described in Section 4.8. This sample path spanned 327.6 seconds with a time step of 0.01

seconds. The estimator used every tenth point of the sample path as a measurement. The FIMLOF estimator for this problem is the same as that derived in Section 6.8. Table 6.9-1 gives the results of the test case.

Table 6.9-1 Results of FIMLOF Fit to Computer Generated Sample Path

	Simulation Value	Estimate
Initial Conditions:		
x_1 (units)	0.00	-1.70 ± 1.3
x_2 (units)	0.00	1.75 ± 1.7
Trend: a (units/sec)	1.00	0.98 ± 0.05
Random Walk Standard Deviation: b (units)	1.00	0.96 ± 1.0
Exp. Corr. Noise:		
Inv. Time Const.: c_1 (1/sec)	0.64	1.76 ± 0.38
Scaling Param.: c_2 (units)	1.56	-0.92 ± 0.22
White Noise Standard Deviation: v (units)	0.10	0.163 ± 0.031

Better results would be obtained for the parameter c_2 by using more data or, if that were not possible, the parameter should be constrained to be non-negative and new results computed for the other parameters. The results show that the estimator does a good job of estimating the trend, random walk, and white noise parameters because there is a great deal of information about these parameters present in the measurements.

The case was simplified by eliminating the exponentially correlated noise state from the simulated sample path. The estimator produced the results contained in Table 6.9-2. Note that the estimates have all improved in quality, with both their uncertainties and the amounts that they differ from the simulation values having decreased.

Table 6.9-2 Results of FIMLOF Fit to Simplified Sample Path

	Simulation Value	Estimate
Initial Condition: x_1 (units)	0.00	-0.12 ± 0.34
Trend: a (units/sec)	1.00	0.98 ± 0.05
Random Walk Standard Deviation: b (units)	1.00	0.98 ± 0.02
White Noise Standard Deviation: v (units)	0.10	0.14 ± 0.008

Chapter 7

Application of Maximum Likelihood Estimation to Fractional Brownian Motion

7.1 Modeling a Non-Markov Process

Because fractional Brownian motion is not a Markov process, it cannot be fully characterized without knowledge of its state for all past time. For this reason, fractional Brownian motion cannot be incorporated into a Kalman filter. These filters propagate from one discrete time to the next storing only the state at the previous time.

The parameters of a non-Markov process may be estimated by using a filter which stores information from every past measurement. One method of doing this, proposed by Lundahl, et. al. [24], takes advantage of the known correlations between the increments of fBm. An estimator of this type may be formulated to use either the measurements or the increments between the measurements as observations. These formulations will be identified as the sum observable and the increment observable respectively.

7.2 Increment Observable Formulation

If $\beta_H(t_k)$ is an fBm process with $\beta_H(0) = 0$, then $E\{\beta_H(t_{k+1})-\beta_H(t_k)\} = 0$ because the fractional Brownian motion process is Gaussian with zero mean.

Equation (5.5-2) shows that if the samples are uniformly spaced ($t_k = k\Delta t$) the autocorrelation function for the increments of the process is [24]

$$\begin{aligned}\phi_{xx}(n) &= E\{[\beta_H(t_{k+n+1})-\beta_H(t_{k+n})][\beta_H(t_{k+1})-\beta_H(t_k)]\} \\ &= \frac{1}{2}\{ E\{[\beta_H(t_{k+n+1}) - \beta_H(t_k)]^2\} \\ &\quad + E\{[\beta_H(t_{k+n}) - \beta_H(t_{k+1})]^2\} \\ &\quad - E\{[\beta_H(t_{k+n+1}) - \beta_H(t_{k+1})]^2\} \\ &\quad - E\{[\beta_H(t_{k+n}) - \beta_H(t_k)]^2\} \} \\ &= \sigma_H^2 \Delta t^{2H}\{|n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H}\}/2 \end{aligned} \quad (7.2\text{-}1)$$

where the parameters σ_H and H are the fractional Brownian motion standard deviation and dimension parameters which were defined in Chapter 5.

Because the increments are zero mean random variables, the correlation between two increments is also equal to their covariance. This allows the increments to be considered to be jointly Gaussian random variables with zero mean and covariance S where the ij^{th} element of S is determined according to

$$S_{ij} = \phi_{xx}(|i-j|) \quad (7.2\text{-}2)$$

The probability density function of the increments then becomes

$$p(z_1,\ldots,z_N) = (2\pi)^{-N/2}\det(S)^{-1/2}\, e^{-(\underline{z}^T S^{-1}\underline{z})/2} \quad (7.2\text{-}3)$$

where

$$\underline{z} = \begin{bmatrix} \beta_H(t_1) - \beta_H(t_0) \\ \vdots \\ \beta_H(t_N) - \beta_H(t_{N-1}) \end{bmatrix} \quad (7.2\text{-}4)$$

Given a set of measurements, maximum likelihood estimation may be used to determine the parameters H and σ_H which maximize the likelihood

that those measurements occurred. The parameters enter the problem through the (N × N) covariance matrix S.

The simple case above may be extended to include more complex systems that are modeled as the sum of several stochastic and/or deterministic processes. Consider a scalar stochastic process $y(t_k)$ which at each discrete time t_k is equal to the sum of a bias $b(t_k)$, several zero mean random variables $x_i(t_k)$ with stationary Gaussian increments, and a stationary white measurement noise $\theta(t_k)$ with $E\{\theta^2\} = \Theta$:

$$y(t_k) = b(t_k) + x_1(t_k) + x_2(t_k) + \ldots + x_N(t_k) + \theta(t_k) \qquad (7.2\text{-}5)$$

The increment from $y(t_k)$ to $y(t_{k+1})$ is equal to

$$\begin{aligned} \Delta y(t_{k+1}) &= y(t_{k+1}) - y(t_k) \\ &= b(t_{k+1}) - b(t_k) + \sum_{i=1}^{N} [x_i(t_{k+1}) - x_i(t_k)] + \theta(t_{k+1}) - \theta(t_k) \end{aligned} \qquad (7.2\text{-}6)$$

Because the stochastic processes are zero mean, the increment can be converted to a zero mean random variable by subtracting the bias term. This gives

$$\Delta y'(t_{k+1}) = \sum_{i=1}^{N} \Delta x_i(t_{k+1}) + \theta(t_{k+1}) - \theta(t_k) \qquad (7.2\text{-}7)$$

where

$$\Delta x_i(t_{k+1}) = x_i(t_{k+1}) - x_i(t_k) \qquad (7.2\text{-}8)$$

The correlation between two increments can now be calculated

$$\begin{aligned} \phi_{yy}(j) &= E\{\Delta y'(t_{k+j+1})\, \Delta y'(t_{k+1})\} \\ &= E\left\{\left[\sum_{i=1}^{N} \Delta x_i(t_{k+j+1}) + \theta(t_{k+j+1}) - \theta(t_{k+j})\right]\right. \\ &\qquad \left.\left[\sum_{m=1}^{N} \Delta x_m(t_{k+1}) + \theta(t_{k+1}) - \theta(t_k)\right]\right\} \end{aligned} \qquad (7.2\text{-}9)$$

If each noise is assumed to be independent of the others, this expression reduces to

$$\phi_{yy}(j) = E\left\{\sum_{i=1}^{N} \Delta x_i(t_{k+j+1})\Delta x_i(t_{k+1})\right\} + E\{[\theta(t_{k+j+1}) - \theta(t_{k+j})][\theta(t_{k+1}) - \theta(t_k)]\}$$

$$= E\left\{\sum_{i=1}^{N} \Delta x_i(t_{k+j+1})\Delta x_i(t_{k+1})\right\} + 2\delta(j)\Theta - \delta(|j|-1)\Theta \qquad (7.2\text{-}10)$$

where $\delta(t) = 1$ at $t = 0$ and is zero elsewhere.

Equation (7.2-12) shows that the autocorrelation function of the increments is equal to the sum of the autocorrelation functions of the increments of the individual stochastic processes. This makes it easy to build $\phi_{yy}(j)$ for a complicated process by summing the correlations of simpler processes. Fractional Brownian motion can be included by simply adding in its autocorrelation which is given by Equation (7.2-1). Also, because $\Delta y'(t_k)$ is a zero mean process, $\phi_{yy}(j)$ is the covariance of its increments. This means that the increments are again characterized by a multivariate Gaussian probability density function.

Let the observable column vector be

$$\underline{z} = \begin{bmatrix} y(t_1) - y(t_o) \\ \cdot \\ \cdot \\ \cdot \\ y(t_N) - y(t_{N-1}) \end{bmatrix} \qquad (7.2\text{-}11)$$

with mean

$$\underline{\mu} = \begin{bmatrix} b(t_1) \\ \cdot \\ \cdot \\ \cdot \\ b(t_N) \end{bmatrix} \qquad (7.2\text{-}12)$$

Then the probability density of the observables is

$$p(z_1,\ldots,z_N;\underline{\alpha}) = (2\pi)^{-N/2}\det[S(\underline{\alpha})]^{-1/2}\, e^{-[(\underline{z}-\underline{\mu}(\underline{\alpha}))^T S(\underline{\alpha})^{-1}(\underline{z}-\underline{\mu}(\underline{\alpha}))]/2} \qquad (7.2\text{-}13)$$

where $\underline{\alpha}$ is the $(m \times 1)$ vector of parameters which characterize the stochastic and deterministic processes which make up the signal of interest and $S(\underline{\alpha})$ is defined by Equation (7.2-2) using the appropriate autocorrelation function. If there is a trend a in the data, $b(t_i) = a\Delta t$. A maximum likelihood estimator can be used to determine the parameters $\underline{\alpha}$ which best match the N data points.

A disadvantage of the increment observable formulation is that it cannot estimate the initial condition associated with a trend. This shortcoming motivates the development of the sum observable formulation.

7.3 Sum Observable Formulation

Recall from Equation (5.4-5) that the nonstationary autocorrelation function for a fractional Brownian motion process is

$$\begin{aligned} \phi_{\beta\beta}(t_j, t_k) &= E\{\beta_H(t_j)\,\beta_H(t_k)\} \\ &= \sigma_H^2\{|t_j|^{2H} + |t_k|^{2H} - |t_j - t_k|^{2H}\}/2 \end{aligned} \qquad (7.3\text{-}1)$$

This autocorrelation function is also the covariance of the zero mean random variables $\beta_H(t_j)$ and $\beta_H(t_k)$. Thus the value of the process at times $t_1, \ldots, t_N$ is a set of jointly Gaussian random variables with zero mean and covariance S where

$$S_{jk} = \phi_{\beta\beta}(t_j, t_k) \qquad (7.3\text{-}2)$$

The sum observable formulation may also be extended to handle processes which are the sum of several deterministic and/or stochastic processes. Consider a scalar stochastic process $y(t_k)$ which at each discrete time t_k is equal to the sum of a bias $b(t_k)$ several zero mean random variables with Gaussian increments $x_i(t_k)$ and a stationary white measurement noise $\theta(t_k)$:

$$y(t_k) = b(t_k) + x_1(t_k) + x_2(t_k) + \ldots + x_N(t_k) + \theta(t_k) \qquad (7.3\text{-}3)$$

If the stochastic processes are zero mean, this observable can be made into a zero mean random variable by subtracting off the bias term leaving

$$y'(t_k) = \sum_{i=1}^{N} x_i(t_k) + \theta(t_k)) \tag{7.3-4}$$

The correlation between two measurements can now be calculated

$$\begin{aligned} \phi_{yy}(t_j,t_k) &= E\{y'(t_j)\, y'(t_k)\} \\ &= E\left\{\left[\sum_{i=1}^{N} x_i(t_j) + \theta(t_j)\right]\left[\sum_{m=1}^{N} x_m(t_k) + \theta(t_k)\right]\right\} \end{aligned} \tag{7.3-5}$$

If each noise is assumed to be independent of the others and $E\{\theta^2\} = \Theta$, this expression can be reduced to

$$\phi_{yy}(t_j,t_k) = E\left\{\sum_{i=1}^{N} x_i(t_j)x_i(t_k)\right\} + \Theta \tag{7.3-6}$$

Equation (7.3-6) shows that the autocorrelation function of the measurements is equal to the sum of the autocorrelation functions of the individual stochastic processes. This makes it easy to build $\phi_{yy}(t_j,t_k)$ for a complicated process by summing the correlations of simpler processes. Fractional Brownian motion can be included by simply adding in its autocorrelation which is given by Equation (7.3-1). Also, because $y'(t_k)$ is a zero mean process, $\phi_{yy}(t_j,t_k)$ is the covariance of the measurements. This means that, just as in the increment observable case, the measurements are characterized by a multivariate Gaussian probability density (7.2-13) with $S(\underline{\alpha})$ being defined by Equation (7.3-2). If there is a trend in the data, $b(t_i) = a_1 + a_2\Delta t_i$.

A disadvantage to using the sum observable joint probability density for maximum likelihood estimation is that the terms on the diagonal of the covariance matrix increase steadily from the upper left to the lower right corner. This tends to make the matrix poorly conditioned as the number of measurements becomes large.

7.4 The Maximum Likelihood Estimator

The fact that both the increment and sum observable formulations reduce to a Gaussian probability density function is very convenient. The maximum likelihood estimator for this function has already been presented in Chapter 6. In fact, the fractional Brownian motion estimator is less difficult to develop analytically because it is a batch estimator and there is no need to assemble the likelihood function recursively using a Kalman filter. Recall that the classical maximum likelihood estimator maximizes the probability density function of the measurements with respect to the unknown parameters. In each of the two formulations developed in the preceding sections, this probability density function is of the form

$$p(z_1,\ldots,z_N;\underline{\alpha}) = (2\pi)^{-N/2}\det[S(\underline{\alpha})]^{-1/2}\, e^{-[(\underline{z}-\underline{\mu}(\underline{\alpha}))^T S(\underline{\alpha})^{-1}(\underline{z}-\underline{\mu}(\underline{\alpha}))]/2} \quad (7.4\text{-}1)$$

where the z_i are the scalar measurements and μ_i their means (subtracted out in computing the covariance $S(\underline{\alpha})$). Following the procedure of Section 6.4, the negative of the logarithm of this function will be minimized in order to simplify the calculations. The negative log-likelihood function without the constant term is thus

$$\zeta(\underline{z};\underline{\alpha}) = \frac{1}{2}\ln\{\det[S(\underline{\alpha})]\} + \frac{1}{2}[\underline{z}-\underline{\mu}(\underline{\alpha})]^T S(\underline{\alpha})^{-1}[\underline{z}-\underline{\mu}(\underline{\alpha})] \quad (7.4\text{-}2)$$

The gradient and an approximation to the Hessian of this function are required so that Newton-Raphson iteration may again be used to perform the minimization. These partial derivatives are identical to those presented in Section 6.6. The gradient vector is

$$\frac{\partial\zeta(\underline{z};\underline{\alpha})}{\partial\alpha_i} = -(\underline{z}-\underline{\mu}(\underline{\alpha}))^T S(\underline{\alpha})^{-1}\frac{\partial\underline{\mu}(\underline{\alpha})}{\partial\alpha_i} - \frac{1}{2}(\underline{z}-\underline{\mu}(\underline{\alpha}))^T S(\underline{\alpha})^{-1}\frac{\partial S(\underline{\alpha})}{\partial\alpha_i} S(\underline{\alpha})^{-1}(\underline{z}-\underline{\mu}(\underline{\alpha}))$$

$$+\frac{1}{2}\mathrm{tr}\left[S(\underline{\alpha})^{-1}\frac{\partial S(\underline{\alpha})}{\partial\alpha_i}\right], \; i = 1, \ldots, m \quad (7.4\text{-}3)$$

and the Fisher information approximation to the Hessian is

$$\frac{\partial^2 \zeta(z;\underline{\alpha})}{\partial\alpha_i \partial\alpha_j} \equiv \mathrm{tr}\left[\frac{\partial \underline{\mu}(\underline{\alpha})}{\partial\alpha_i} \frac{\partial \underline{\mu}(\underline{\alpha})^T}{\partial\alpha_i} S(\underline{\alpha})^{-1} + \frac{1}{2} \frac{\partial S(\underline{\alpha})}{\partial\alpha_i} S(\underline{\alpha})^{-1} \frac{\partial S(\underline{\alpha})}{\partial\alpha_j} S(\underline{\alpha})^{-1} \right] \quad (7.4\text{-}4)$$

$$i = 1, \ldots, m; \quad j = 1, \ldots, m$$

The maximum likelihood estimator is implemented by fixing the unknown parameters, calculating the likelihood function and its partial derivatives, performing a Newton-Raphson update to the parameters, as described in Section 6.5, and repeating the process until the parameters converge. A drawback to this method is the computational burden that comes with inverting the covariance matrix S, which has as many rows and columns as there are measurements.

7.5 Estimation Results with Pure Fractional Brownian Motion

7.5.1 Increment Observable Model

Given scalar measurements $y(t_i)$ which are taken at fixed intervals of Δt from an fBm signal, increment and sum observable estimators may be developed to estimate the parameters σ_H and H. For the increment observable formulation the ij^{th} element of the covariance matrix is given by

$$S_{ij} = \sigma_H^2 \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \}/2, \quad n = |i-j| \quad (7.5.1\text{-}1)$$

The partial derivatives of these elements with respect to the parameters are

$$\frac{\partial S_{ij}}{\partial \sigma_H} = \sigma_H \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \} \quad (7.5.1\text{-}2)$$

$$\frac{\partial S_{ij}}{\partial H} = \sigma_H^2 \Delta t^{2H} \{ \ln|n+1|\,|n+1|^{2H} - 2\ln|n|\,|n|^{2H} + \ln|n-1|\,|n-1|^{2H} \}$$

$$+ \sigma_H^2 (\ln \Delta t) \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \} \quad (7.5.1\text{-}3)$$

where the negative logarithm of a zero argument is ignored because it is multiplied by zero. The measurement vector is

$$\underline{z} = \begin{bmatrix} y(t_1) - y(t_o) \\ \cdot \\ \cdot \\ \cdot \\ y(t_N) - y(t_{N-1}) \end{bmatrix} \qquad (7.5.1\text{-}4)$$

The mean is $\underline{\mu} = 0$ so that its partial derivatives are

$$\frac{\partial \underline{\mu}}{\partial \sigma_H} = 0 \qquad (7.5.1\text{-}5)$$

$$\frac{\partial \underline{\mu}}{\partial H} = 0 \qquad (7.5.1\text{-}6)$$

7.5.2 Sum Observable Formulation

For the sum observable formulation, the covariance matrix is defined by

$$S_{ij} = \sigma_H^2 \{ |t_i|^{2H} + |t_j|^{2H} - |t_i - t_j|^{2H} \}/2 \qquad (7.5.2\text{-}1)$$

with partial derivatives

$$\frac{\partial S_{ij}}{\partial \sigma_H} = \sigma_H \{ |t_i|^{2H} + |t_j|^{2H} - |t_i - t_j|^{2H} \} \qquad (7.5.2\text{-}2)$$

$$\frac{\partial S_{ij}}{\partial H} = \sigma_H^2 \{ \ln|t_i|\,|t_i|^{2H} + \ln|t_j|\,|t_j|^{2H} - \ln|t_i - t_j|\,|t_i - t_j|^{2H} \} \qquad (7.5.2\text{-}3)$$

The measurement vector is

$$\underline{z} = \begin{bmatrix} y(t_1) \\ \cdot \\ \cdot \\ \cdot \\ y(t_N) \end{bmatrix} \qquad (7.5.2\text{-}4)$$

and partial derivatives of the mean $\underline{\mu}$ are zero

$$\frac{\partial \mu}{\partial \sigma_H} = 0 \qquad (7.5.2\text{-}5)$$

$$\frac{\partial \mu}{\partial H} = 0 \qquad (7.5.2\text{-}6)$$

7.5.3 Fits to Computer Generated Sample Paths

The above estimators were implemented to determine the parameters of a sample path which was generated using the technique described in Section 5.8.2. The sample paths were made up of 128 points spaced one second apart. The parameter values used in the simulation are given in Table 7.5-1 along with the results of the fits. In this table, the "±" value is the Cramer-Rao lower bound described in Section 2.4. These results show that the estimators produce identical results, and that they are capable of determining the fBm parameters with accuracy on the order of the Cramer-Rao lower bound. This accuracy may be improved by increasing the number of measurements. Because the increment observable method is better suited for implementation, the remaining examples of this chapter will use that formulation.

Table 7.5-1 Parameter Estimates from Fits To Pure fBm

	Simulation Value	INCREMENT	SUM
Standard Deviation: σ_H	1.0	0.988 ± 0.07	0.988 ± 0.07
Dimension: H	0.1	0.109 ± 0.03	0.109 ± 0.03

7.6 Results with Fractional Brownian Motion Plus a Linear Trend

If a linear trend $a\Delta t$ is summed with the fBm signal described in Section 7.5.1, the estimator changes very little. The increment observable formulation of the ij^{th} element of the covariance matrix is still given by

$$S_{ij} = \sigma_H^2 \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \}/2, \quad n = |i-j| \tag{7.6-1}$$

The partial derivatives of these elements with respect to the parameters are

$$\frac{\partial S_{ij}}{\partial a} = 0 \tag{7.6-2}$$

$$\frac{\partial S_{ij}}{\partial \sigma_H} = \sigma_H \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \} \tag{7.6-3}$$

$$\frac{\partial S_{ij}}{\partial H} = \sigma_H^2 \Delta t^{2H} \{ \ln|n+1|\,|n+1|^{2H} - 2\ln|n|\,|n|^{2H} + \ln|n-1|\,|n-1|^{2H} \}$$

$$+ \sigma_H^2 (\ln \Delta t) \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \} \tag{7.6-4}$$

where the parameter a defines the slope of the trend. The measurement vector becomes

$$\underline{z} = \begin{bmatrix} y(t_1) - y(t_o) \\ \cdot \\ \cdot \\ \cdot \\ y(t_N) - y(t_{N-1}) \end{bmatrix} \tag{7.6-5}$$

The mean and its partial derivatives are

$$\underline{\mu} = \begin{bmatrix} a\Delta t \\ \cdot \\ \cdot \\ \cdot \\ a\Delta t \end{bmatrix} \tag{7.6-6}$$

$$\frac{\partial \underline{\mu}}{\partial a} = \begin{bmatrix} \Delta t \\ \cdot \\ \cdot \\ \cdot \\ \Delta t \end{bmatrix} \tag{7.6-7}$$

$$\frac{\partial \underline{\mu}}{\partial \sigma_H} = 0 \tag{7.6-8}$$

$$\frac{\partial \underline{\mu}}{\partial H} = 0 \tag{7.6-9}$$

A similar estimator may be developed using the sum observable formulation.

Table 7.6-1 contains results of fits to a sample path which consists of fBm with an additive linear trend. In the simulation, the fBm noise parameters were $\sigma_H = 0.7$ and $H = 0.4$. A single fBm sample path was generated and then the two cases were created by adding different trends to that sample path. These results show that the magnitude of the trend did not affect the estimate of the fBm parameters. It is also important to note that the initial condition of the trend cannot be estimated using the increment observable formulation.

Table 7.6-1 Parameter Estimates from Fit to fBm Plus Trend (σ_H=.7, H=.4)

Case:	a	σ_H	H
1. a = 1.5	1.486 ± 0.02	0.914 ± 0.04	0.301 ± 0.03
2. a = 5.0	4.986 ± 0.02	0.914 ± 0.04	0.301 ± 0.03

7.7 Results with fBm Plus White Measurement Noise

Addition of white Gaussian measurement noise with zero mean and standard deviation σ_m to a fractional Brownian motion signal forces a modification to the covariance matrix. In the increment observable case, the ij^{th} element of this covariance matrix is given by

$$S_{ij} = \sigma_H^2 \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \}/2$$

$$+ 2\sigma_m^2 \delta(n) - \sigma_m^2 \delta(|n|-1) , \quad n = |i-j| \qquad (7.7\text{-}1)$$

where $\delta(n) = 1$ if $n = 0$ and 0 elsewhere. The partial derivatives of these elements with respect to the parameters are

$$\frac{\partial S_{ij}}{\partial \sigma_H} = \sigma_H \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \} \qquad (7.7\text{-}2)$$

$$\frac{\partial S_{ij}}{\partial H} = \sigma_H^2 \Delta t^{2H} \{ \ln|n+1|\,|n+1|^{2H} - 2\ln|n|\,|n|^{2H} + \ln|n-1|\,|n-1|^{2H} \}$$

$$+ \sigma_H^2 (\ln \Delta t) \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \} \qquad (7.7\text{-}3)$$

$$\frac{\partial S_{ij}}{\partial \sigma_m} = \begin{cases} 4\sigma_m & \text{if } n = 0 \\ -2\sigma_m & \text{if } n = \pm 1 \\ 0 & \text{else} \end{cases} \qquad (7.7\text{-}4)$$

Parameters were fit to three separate sample paths using the estimator above. These cases had fBm parameters $\sigma_H = 0.7$ and $H = 0.4$ and varying levels of measurement noise. The resulting parameter estimates are shown in Table 7.7-1. For σ_m of 0.025 and 0.25, the parameter estimates converged, but when σ_m was set to 1.0, the measurement noise parameter diverged.

There are two possible reasons for the failure of the algorithm. It is possible that with only 128 measurements, the likelihood function was not sharply peaked. This can be seen by examining the singular value decomposition of the symmetric Fisher information matrix which may be written as

$$I_{ij} = \sum_{i=1}^{q} \sigma_i \underline{v}_i \underline{v}_i^T \qquad (7.7\text{-}5)$$

where q is the number of parameters, σ_i is the i^{th} singular value, and $\underline{v}_i$ is the direction associated with the i^{th} singular value. The parameter updates are then given by

$$\Delta\underline{\alpha} = \sum_{i=1}^{q} \frac{1}{\sigma_i} \underline{v}_i \underline{v}_i^T B \qquad (7.7\text{-}6)$$

where B is the negative gradient vector. Because the Fisher information matrix will have small singular values corresponding to directions in which there is little information, the corrections $\Delta\underline{\alpha}$ will be large in these directions [15]. This would send the corrections far past the optimal parameter estimates during the update. If the corrections are large enough, the parameter estimates are no longer close to the optimal values and the code may converge to an extraneous local minimum.

It is also possible that for 128 measurements, the Fisher information matrix was not a valid approximation to the Hessian. Either problem may be solved by the incorporation of more measurements thus increasing the amount of information used by the estimator and improving the Fisher information approximation to the Hessian. Another fit was performed using 200 points, and the algorithm converged to the estimates listed in Table 7.7-1.

For comparison, another fit to the case with $\sigma_m = 0.25$ was also performed. This fit demonstrates the fact that the Cramer-Rao lower bound will decrease when additional measurements are included. This does not assure a better solution, however, as seen from the results in the table.

Table 7.7-1 Parameter Estimates from Fit to fBm Plus Measurement Noise ($\sigma_H = .7$, $H = .4$)

Case:	σ_m	σ_H	H
1. σ_m = 0.025 (128 pts)	0.271 ± 0.09	0.522 ± 0.09	0.599 ± 0.11
2. σ_m = 0.25 (128 pts)	0.353 ± 0.16	0.651 ± 0.17	0.461 ± 0.13
3. σ_m = 0.25 (200 pts)	0.402 ± 0.08	0.571 ± 0.10	0.556 ± 0.10
4. σ_m = 1.00 (128 pts)	---	---	---
5. σ_m = 1.00 (200 pts)	0.838 ± 0.19	0.899 ± 0.32	0.350 ± 0.13

7.8 Results with fBm Plus Exponentially Correlated Noise

Let $x_1(t_k)$ be a discrete fBm process with parameters σ_H and H, and let $x_2(t_k)$ be a discrete exponentially correlated noise process with reciprocal time constant c_1 and scaling parameter c_2. The autocorrelation function for $x_2(t_k)$ is then given by

$$E\{x_2(t_k)x_2(t_{k+n})\} = \frac{1}{2} c_1 c_2^2 e^{-c_1 |n| \Delta t} \qquad (7.8\text{-}1)$$

Let $y(t_k)$ be equal to the sum of the two independent processes and let $\Delta y(t_k)$ be its increments. By Equation (7.2-11), the autocorrelation function of the increments is given by

$$\phi_{xx}(n) = E\{\Delta x_1(t_k)\Delta x_1(t_{k+n})\} + E\{\Delta x_2(t_k)\Delta x_2(t_{k+n})\} \qquad (7.8\text{-}2)$$

While the first term of Equation (7.8-2) is the known autocorrelation of fBm increments, the second term may be manipulated to get a more convenient form.

$$\phi_{xx}(n) = E\{\Delta x_1(t_k)\Delta x_1(t_{k+n})\} + E\{[x_2(t_k) - x_2(t_{k-1})][x_2(t_{k+n}) - x_2(t_{k+n-1})]\}$$

$$= E\{\Delta x_1(t_k)\Delta x_1(t_{k+n})\} + E\{x_2(t_k)x_2(t_{k+n})\} - E\{x_2(t_k)x_2(t_{k+n-1})\}$$

$$- E\{x_2(t_{k-1})x_2(t_{k+n})\} + \{x_2(t_{k-1})x_2(t_{k+n-1})\} \quad (7.8\text{-}3)$$

Using Equations (7.2-1) and (7.8-1) this reduces to

$$\phi_{xx}(n) = \sigma_H^2 \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \}/2$$

$$+ \frac{1}{2} c_1 c_2^2 [2e^{-c_1|n|\Delta t} - e^{-c_1|n-1|\Delta t} - e^{-c_1|n+1|\Delta t}] \quad (7.8\text{-}4)$$

Because both processes are unbiased, the autocorrelation function for the increments is equal to their covariance so that

$$S_{ij} = \phi_{xx}(n) , \quad n = |i-j| \quad (7.8\text{-}5)$$

The partial derivatives of S with respect to the parameters are given by

$$\frac{\partial S_{ij}}{\partial \sigma_H} = \sigma_H \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \} \quad (7.8\text{-}6)$$

$$\frac{\partial S_{ij}}{\partial H} = \sigma_H^2 \Delta t^{2H} \{ \ln|n+1|\,|n+1|^{2H} - 2\ln|n|\,|n|^{2H} + \ln|n-1|\,|n-1|^{2H} \}$$

$$+ \sigma_H^2 (\ln \Delta t) \Delta t^{2H} \{ |n+1|^{2H} - 2|n|^{2H} + |n-1|^{2H} \} \quad (7.8\text{-}7)$$

$$\frac{\partial S_{ij}}{\partial c_1} = \frac{1}{2} c_2^2 [2e^{-c_1|n|\Delta t} - e^{-c_1|n-1|\Delta t} - e^{-c_1|n+1|\Delta t}]$$

$$- \frac{1}{2} \Delta t\, c_1 c_2^2 [2|n|e^{-c_1|n|\Delta t} - |n-1|e^{-c_1|n-1|\Delta t} - |n+1|e^{-c_1|n+1|\Delta t}] \quad (7.8\text{-}8)$$

$$\frac{\partial S_{ij}}{\partial c_2} = c_1 c_2 [2e^{-c_1|n|\Delta t} - e^{-c_1|n-1|\Delta t} - e^{-c_1|n+1|\Delta t}] \quad (7.8\text{-}9)$$

The measurement vector and the partial derivatives of the mean vector are are

$$\underline{z} = \begin{bmatrix} y(t_1) - y(t_0) \\ \vdots \\ y(t_N) - y(t_{N-1}) \end{bmatrix} \quad (7.8\text{-}10)$$

$$\frac{\partial \mu}{\partial \sigma_H} = 0 \qquad (7.8\text{-}11)$$

$$\frac{\partial \mu}{\partial H} = 0 \qquad (7.8\text{-}12)$$

$$\frac{\partial \mu}{\partial c_1} = 0 \qquad (7.8\text{-}13)$$

$$\frac{\partial \mu}{\partial c_2} = 0 \qquad (7.8\text{-}14)$$

A fit to a computer generated sample path with parameters $\sigma_H = 0.7$, $H = 0.4$, $c_1 = 0.5$, and $c_2 = 0.9$, failed. The fit used 128 measurements spaced at one second intervals. This example demonstrates the fact that the addition of a fourth parameter greatly increases the computational burden while also requiring many more measurements to get a reasonable estimate. For comparison, using 128 measurements, one iteration of the case involving fBm plus white measurement noise required 44.3 seconds of CPU time, while a single iteration of the case of fBm plus exponentially correlated noise took 72.0 seconds to complete. Of course, the CPU time required to accomplish an iteration depends on the complexity of the parameter partial derivatives as well as on their number.

Increasing the number of measurements by even a small amount also greatly increases the time required to perform an iteration because the operations required to invert the covariance matrix increases in proportion to the number of measurements cubed. For example, in the exponentially correlated noise case, changing the number of measurements from 128 to 200 resulted in a corresponding increase of CPU time from 72.0 to 236.7 seconds.

7.9 Results of Fit to Accelerometer Data

The maximum likelihood estimator was used to perform a fit of the fractional Brownian motion model to the accelerometer data described in Section 6.8. In order to decrease the number of data points, while still

capturing the low frequency behavior of the system, the 1 Hz data was averaged over intervals of 150 seconds. The fit was then performed using 128 of these averaged measurements. The averaging process had the effect of limiting the frequency domain information contained in the signal to the range between 5.2×10^{-5} Hz and 3.33×10^{-3} Hz. The PSD of the averaged data was nearly identical to that shown in Figure 6.8-2 over the frequency range of interest. The results of this successful fit are contained in Table 7.9-1. A value of H = 0.212 is indicative of a process with log-log PSD slope equal to -1.424, while the PSD slope of Figure 6.8-2 actually displays a -1 log-log PSD slope.

Table 7.9-1 Parameter Estimates from Fit to Accelerometer Data

Parameter:	Estimate	Cramer-Rao Lower Bound
σ_H	2.725	0.72
H	0.212	0.04

Appendix A

Review of Frequency Domain Properties of Convolution and Correlation

Chapter 3 employs a frequency domain property of autocorrelation to determine an estimate of the power spectral density of a stationary, ergodic, stochastic process. Correlation is very similar to the more well known operation of convolution. This appendix reviews the frequency domain properties of these operations.

The convolution $(x*y)(t) = x(t)*y(t)$ of two time domain functions $x(t)$ and $y(t)$ with Fourier transforms $X(f)$ and $Y(f)$ is defined to be [34]

$$x(t)*y(t) \equiv \int_{-\infty}^{\infty} x(\tau)\, y(t-\tau)\, d\tau \qquad \text{(A-1)}$$

$$= \int_{-\infty}^{\infty} x(t-\tau')\, y(\tau')\, d\tau' = y(t)*x(t)$$

where the second line is otained by a change of integration variable, so that convolution is a commutative operation.

The Fourier transform of Equation (A-1) is

$$\int_{-\infty}^{\infty} x(t)*y(t)\, e^{-2\pi i f t}\, dt = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x(\tau)\, y(t-\tau)\, d\tau\right] e^{-2\pi i f t}\, dt \qquad \text{(A-2)}$$

where the frequency f is given in Hz.

Making the change of variables $t = \tau + z$ leads to

$$\int_{-\infty}^{\infty} x(t)*y(t)\, e^{-2\pi i f t}\, dt = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(\tau)\, y(z)\, e^{-2\pi i f(\tau+z)}\, d\tau\, dz$$

$$= \int_{-\infty}^{\infty} x(\tau)\, e^{-2\pi i f \tau}\, d\tau \int_{-\infty}^{\infty} y(z)\, e^{-2\pi i f z}\, dz$$

$$= X(f)\, Y(f) \qquad \text{(A-3)}$$

Thus the Fourier transform of a convolution is equal to the product of the individual Fourier transforms of the functions of interest. Equation (A-3) is known as the convolution theorem [34].

In the case of correlation, the results are slightly different from that of convolution. The correlation of two time domain functions x(t) and y(t) is defined to be [8]

$$\phi_{xy}(\tau) \equiv \int_{-\infty}^{\infty} x(t+\tau)\, y(t)\, dt$$

$$= \int_{-\infty}^{\infty} x(t+\tfrac{\tau}{2})\, y(t-\tfrac{\tau}{2})\, dt \qquad \text{(A-4)}$$

where the second line is obtained by a change of variable.

Taking the Fourier transform of both sides of Equation (A-4) gives

$$\int_{-\infty}^{\infty} \phi_{xy}(\tau)\, e^{-2\pi i f \tau}\, d\tau = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x(t+\tau)\, y(t)\, dt\right] e^{-2\pi i f \tau}\, d\tau \qquad \text{(A-5)}$$

Making the change of variables $\tau = -t + z$ leads to

$$\int_{-\infty}^{\infty} \phi_{xy}(\tau)\, e^{-2\pi i f \tau}\, d\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(z)\, y(t)\, e^{-2\pi i f(-t+z)}\, dt\, dz$$

$$= \int_{-\infty}^{\infty} x(z)\, e^{-2\pi i f z}\, dz \int_{-\infty}^{\infty} y(t)\, e^{2\pi i f t}\, dt$$

$$= X(f)\, Y^*(f) \qquad \text{(A-6)}$$

where the superscript * denotes complex conjugate.

Equation (A-6) shows that the frequency domain properties of a correlation are slightly different from those of a convolution. The Fourier transform of a correlation is equal to the product of the Fourier transform of the first term with the complex conjugate of the Fourier transform of the second.

Appendix B

Derivative of the Natural Logarithm of a Determinant

For an n×n matrix S the determinant is [1]

$$\det(S) = \sum_{i_1 \ldots i_n} \varepsilon_{i_1 \ldots i_n} S_{1 i_1} \ldots S_{n i_n} \tag{B-1}$$

$$= \sum_{i_1 \ldots i_n} \varepsilon_{i_1 \ldots i_n} S_{i_1 1} \ldots S_{i_n n} \tag{B-1}'$$

where S_{ij} denotes the ij^{th} element of S and

$$\varepsilon_{i_1 \ldots i_n} = \begin{cases} 0 & \text{if any } i_j = i_k \ (j \neq k) \\ +1 & \text{if } i_1 \ldots i_n \text{ is an even permutation of } 1 \ldots n \\ -1 & \text{if } i_1 \ldots i_n \text{ is an odd permutation of } 1 \ldots n \end{cases} \tag{B-2}$$

For any two square matrices A and B we have [1]

$$\det(AB) = \det(A) \det(B) \tag{B-3}$$

The inverse of S is [1]

$$S^{-1} = \frac{1}{\det(S)} (\text{matrix of cofactors of } S)^T \tag{B-4}$$

where superscript T denotes transpose and the entries in the matrix of cofactors of S are

$$(\text{cofactor } S)_{ab} = (-1)^{a+b} \det (\text{S with row a and column b deleted}) \tag{B-5}$$

Note that if S is symmetric, then

$$S = S^T, \quad S^{-1} = (S^{-1})^T \tag{B-6}$$

Differentiating Equation (B-1) with respect to α = some α_i, we obtain [1]

$$\frac{\partial \det(S)}{\partial \alpha} = \sum_{i=1}^{n} \det(\text{ S with row i replaced by its derivative })$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} (\text{cofactor S})_{ij} \frac{\partial S_{ij}}{\partial \alpha} \tag{B-7}$$

If S is symmetric this implies

$$\frac{\partial \det(S)}{\partial \alpha} = \det(S)\ \text{trace}\left[S^{-1}\frac{\partial S}{\partial \alpha}\right] \tag{B-8}$$

where the trace of a square matrix is the sum of the diagonal terms. Therefore (see Equation (6.6-2))

$$\frac{\partial \ln[\det(S)]}{\partial \alpha} = \text{trace}\left[S^{-1}\frac{\partial S}{\partial \alpha}\right] \tag{B-9}$$

Appendix C

Derivation of Equation (6.6-4)

This appendix fills in some of the details which were left out of the same derivation in Reference [37]. To derive Equation (6.6-4) it is necessary to know the third and fourth moments of a multivariate Gaussian distribution. These moments are derived first. Let $\underline{c}$ be a nonrandom vector, let A and B be symmetric, nonrandom matrices, and let $\underline{r}$ be a multivariate, zero mean, Gaussian random variable with covariance matrix S. The third moment of $\underline{r}$ is given by

$$E\{ [\underline{c}^T\underline{r}] [\underline{r}^T A\underline{r}] \} = \sum_{i,j,k} c_i A_{jk} E\{ r_i r_j r_k \} \qquad \text{(C-1)}$$

Now

$$E\{ r_i r_j r_k \} = \frac{1}{\sqrt{(2\pi)^n \det(S)}} \int \cdots \int r_i r_j r_k \; e^{-\underline{r}^T S^{-1}\underline{r}/2} \qquad \text{(C-2)}$$

where $\underline{r}$ has n elements. Let $\underline{y} = B\underline{r}$ be such that its covariance, given by cov($\underline{y}$), is a diagonal matrix with ones on the diagonal:

$$\delta_{jk} = [\text{cov}(\underline{y})]_{jk} = \sum_{a,b} B_{ja} B_{kb} S_{ab} \qquad \text{(C-3)}$$

Since $\underline{r} = D\underline{y}$ with $D = B^{-1}$,

$$S_{jk} = \sum_{a} D_{ja} D_{ka} \qquad \text{(C-4)}$$

Now consider $E\{ y_i y_j y_k \}$ given by an integral like Equation (C-2) with S diagonal. If all the i,j,k are different, the integral is zero (mean of a first order Gaussian). If two indices are the same and one different the integral is again zero for the same reason applied to the one different index. Finally, if all three indices are the same, the integral is zero as the third moment of a first order Gaussian [18]. Since $\underline{r} = D\underline{y}$, $E\{ r_i r_j r_k \} = 0$ and Equation (C-1) is zero.

The fourth moment of a scalar Gaussian random variable is given by [18]

$$E\{y_i^4\} = 3\, \mathrm{cov}(y_i)^2 \tag{C-5}$$

This implies

$$E\{ y_i y_j y_k y_m \} = \begin{cases} 3 & \text{if } i = j = k = m \\ 1 & \text{if } i = j \neq k = m \\ 0 & \text{if some index is different from others} \end{cases} \tag{C-6}$$

Substituting the known relation for r_i gives

$$E\{ r_a r_b r_c r_d \} = E\{ \sum_i D_{ai} y_i \sum_j D_{aj} y_j \sum_k D_{ak} y_k \sum_m D_{am} y_m \}$$

$$= 3 \sum_i D_{ai}D_{bi}D_{ci}D_{di} + \sum_i \sum_{k \neq i} D_{ai}D_{bi}D_{ck}D_{dk}$$

$$+ \sum_i \sum_{k \neq i} D_{ai}D_{bk}D_{ci}D_{dk} + \sum_i \sum_{k \neq i} D_{ai}D_{bk}D_{ck}D_{di}$$

$$= S_{ab}S_{cd} + S_{ac}S_{bd} + S_{ad}S_{bc} \tag{C-7}$$

Equation (C-7) may then be generalized to

$$E\{ [\underline{r}^T A \underline{r}\,] [\underline{r}^T B \underline{r}\,] \} = E\{ \sum_{a,b} A_{ab} r_a r_b \sum_{c,d} B_{cd} r_c r_d \}$$

$$= \sum_{a,b,c,d} A_{ab}B_{cd}\, E\{ r_a r_b r_c r_d \}$$

$$= \sum_{a,b,c,d} A_{ab}B_{cd}\,[S_{ab}S_{cd} + S_{ac}S_{bd} + S_{ad}S_{bc}]$$

$$= [\sum_{a,b} A_{ab}S_{ab}\,]\,[\sum_{c,d} B_{cd}S_{cd}\,] + 2 \sum_{a,b,c,d} A_{ab}B_{cd}S_{ac}S_{bd}$$

$$= [\mathrm{tr}(SA)]\,[\mathrm{tr}(SB)] + 2\,\mathrm{tr}[SASB] \qquad \text{(C-8)}$$

The fact that the third moment is zero and Equation (C-8) are now used to derive Equation (6.6-4) from Equation (6.6-3). The explicit dependence of $\underline{r}$ and S on t_k and $\underline{\alpha}$ will be left out on the right hand side for clarity. Inserting Equation (6.6-2) into the term within the summation in Equation (6.6-3) gives the following

$$\begin{aligned}
E\left\{ \frac{\partial \zeta(\underline{z}(t_k)\,|\,\underline{z}^{k-1};\underline{\alpha})}{\partial \alpha_i} \frac{\partial \zeta(\underline{z}(t_k)\,|\,\underline{z}^{k-1};\underline{\alpha})}{\partial \alpha_j} \right\} = E\Bigg\{ & \left[\underline{r}^T S^{-1}\frac{\partial \underline{r}}{\partial \alpha_i}\right]\left[\underline{r}^T S^{-1}\frac{\partial \underline{r}}{\partial \alpha_j}\right] \\
& + \frac{1}{4}\left[\underline{r}^T S^{-1}\frac{\partial S}{\partial \alpha_i} S^{-1}\underline{r}\right]\left[\underline{r}^T S^{-1}\frac{\partial S}{\partial \alpha_j} S^{-1}\underline{r}\right] \\
& + \frac{1}{4}\,\mathrm{tr}\left[S^{-1}\frac{\partial S}{\partial \alpha_i}\right]\mathrm{tr}\left[S^{-1}\frac{\partial S}{\partial \alpha_j}\right] \\
& - \frac{1}{2}\left[\underline{r}^T S^{-1}\frac{\partial \underline{r}}{\partial \alpha_i}\right]\left[\underline{r}^T S^{-1}\frac{\partial S}{\partial \alpha_j} S^{-1}\underline{r}\right] \\
& - \frac{1}{2}\left[\underline{r}^T S^{-1}\frac{\partial S}{\partial \alpha_i} S^{-1}\underline{r}\right]\left[\underline{r}^T S^{-1}\frac{\partial \underline{r}}{\partial \alpha_j}\right] \\
& + \frac{1}{2}\,\mathrm{tr}\left[S^{-1}\frac{\partial S}{\partial \alpha_i}\right]\left[\underline{r}^T S^{-1}\frac{\partial \underline{r}}{\partial \alpha_j}\right] \\
& + \frac{1}{2}\left[\underline{r}^T S^{-1}\frac{\partial \underline{r}}{\partial \alpha_i}\right]\mathrm{tr}\left[S^{-1}\frac{\partial S}{\partial \alpha_j}\right] \\
& - \frac{1}{4}\,\mathrm{tr}\left[S^{-1}\frac{\partial S}{\partial \alpha_i}\right]\left[\underline{r}^T S^{-1}\frac{\partial S}{\partial \alpha_j} S^{-1}\underline{r}\right] \\
& - \frac{1}{4}\left[\underline{r}^T S^{-1}\frac{\partial S}{\partial \alpha_i} S^{-1}\underline{r}\right]\mathrm{tr}\left[S^{-1}\frac{\partial S}{\partial \alpha_j}\right]\Bigg\} \qquad \text{(C-9)}
\end{aligned}$$

At this point, recall that the covariance matrix S is a nonrandom function and note that the partial derivative of the residual $\underline{r}$ with respect to the unknown parameters is also a nonrandom function. Using these facts

along with the moments derived in this appendix eliminates the 4th, 5th, 6th, and 7th terms of Equation (C-9) and leaves

$$E\left\{ \frac{\partial \zeta(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha})}{\partial \alpha_i} \frac{\partial \zeta(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha})}{\partial \alpha_j} \right\} = \mathrm{tr}\left[\frac{\partial \underline{r}^T}{\partial \alpha_i} \frac{\partial \underline{r}}{\partial \alpha_i} S^{-1} \right]$$
$$+ \frac{1}{4} \left\{ \mathrm{tr}\left[\frac{\partial S}{\partial \alpha_i} S^{-1} \right] \mathrm{tr}\left[\frac{\partial S}{\partial \alpha_j} S^{-1} \right] \right.$$
$$\left. + 2\,\mathrm{tr}\left[\frac{\partial S}{\partial \alpha_i} S^{-1} \frac{\partial S}{\partial \alpha_j} S^{-1} \right] \right\}$$
$$+ \frac{1}{4} \mathrm{tr}\left[S^{-1} \frac{\partial S}{\partial \alpha_i} \right] \mathrm{tr}\left[S^{-1} \frac{\partial S}{\partial \alpha_j} \right]$$
$$- \frac{1}{4} \mathrm{tr}\left[S^{-1} \frac{\partial S}{\partial \alpha_i} \right] \mathrm{tr}\left[S^{-1} \frac{\partial S}{\partial \alpha_j} \right]$$
$$- \frac{1}{4} \mathrm{tr}\left[S^{-1} \frac{\partial S}{\partial \alpha_i} \right] \mathrm{tr}\left[S^{-1} \frac{\partial S}{\partial \alpha_j} \right] \qquad \text{(C-10)}$$

Thus

$$E\left\{ \frac{\partial^2 \zeta(\underline{z}(t_k) \mid \underline{z}^{k-1};\underline{\alpha})}{\partial \alpha_i \partial \alpha_j} \right\} = \mathrm{tr}\left[\frac{\partial \underline{r}^T}{\partial \alpha_i} \frac{\partial \underline{r}}{\partial \alpha_j} S^{-1} \right] + \frac{1}{2} \mathrm{tr}\left[\frac{\partial S}{\partial \alpha_i} S^{-1} \frac{\partial S}{\partial \alpha_j} S^{-1} \right] \qquad \text{(C-11)}$$

which becomes Equation (6.6-4) when the summation and the dependence on t_k and $\underline{\alpha}$ are reapplied:

$$A_{ij} \equiv \sum_{k=0}^{N} \mathrm{tr}\left[\frac{\partial \underline{r}(t_k;\underline{\alpha})}{\partial \alpha_i} \frac{\partial \underline{r}(t_k;\underline{\alpha})^T}{\partial \alpha_j} S(t_k;\underline{\alpha})^{-1} \right.$$
$$\left. + \frac{1}{2} \frac{\partial S(t_k;\underline{\alpha})}{\partial \alpha_i} S(t_k;\underline{\alpha})^{-1} \frac{\partial S(t_k;\underline{\alpha})}{\partial \alpha_j} S(t_k;\underline{\alpha})^{-1} \right] \qquad \text{(C-12)}$$

Appendix D

Numerical Algorithms

D.1 FORTRAN Storage Convention

An m×n matrix $Z = (Z_{ij})$ is stored as a vector in FORTRAN with the first (row) index being the most rapidly varying (the opposite of the convention in PL/I, Pascal, and other languages):

$$Z_{ij} = Z[\, j*m+i \,], \quad 1 \le i \le m,\ 1 \le j \le n$$

This FORTRAN storage convention is taken advantage of in the software written to do the data analyses discussed in this thesis, when part of a multidimensional array W(m,n,k) is sent to a subroutine. If the subroutine expects W(m,n), and the j^{th} m×n part of W is to be sent to the subroutine, the calling program sends the address of W(1,1,j) to the subroutine. Some computer languages would not allow a mismatching of dimensions between a calling and called subroutine, but FORTRAN does, and advantage is taken of the fact.

D.2 Symmetric Matrix Manipulation

An n×n symmetric matrix $S = (S_{ij})$ with $S_{ij} = S_{ji}$ is stored in lower diagonal form as a vector in the FORTRAN software. Thus, the FORTRAN vector is

$$S_{ij} = S[\, (j*(j-1))+i \,], \quad 1 \le i \le j \le n$$

Use of lower diagonal form saves about a factor of 2 in computation time and storage, which can be important when dealing with very large matrices.

There are subroutines to multiply non-symmetric and symmetric matrices, with appropriate handling of the indices in each case. If a symmetric product results from some matrix multiplication combination, such as ZSZ^T, then there is one subroutine to do the whole combination, with non-symmetric Z and symmetric S as input, and symmetric result as output.

Of great importance in estimation algorithms is the inversion of a symmetric matrix A (such as the Fisher information matrix) and the possible simultaneous solution for a set of linear equations with the given matrix as a coefficient matrix (for determining adjustments in a Newton-Raphson iteration). A private subroutine SYMINV written by Norman Brenner in 1968 was used to perform this operation. SYMINV takes as input a symmetric matrix A stored in lower diagonal form, a right hand side vector B, and simultaneously in place inverts the matrix A and solves for the vector X in the equation $AX = B$. The operations being done in place means that the symmetric matrix A^{-1} and solution vector X replace A and B, respectively, at the end of the subroutine execution.

Gaussian elimination is used in SYMINV. In order to have the algorithm work with the lower diagonal storage, interchange of rows and columns is not used, and the Gaussian elimination pivots are chosen to be diagonal elements.

The only reason interchange of rows and columns is used in Gaussian elimination is to allow off-diagonal pivots. At a given stage of the algorithm execution, the next pivot should be the largest element remaining in the matrix. However, for the covariance type matrices dealt with in estimation theory, the largest elements are likely to be on the diagonal, so it is no restriction not to allow row and column interchange and only choose the pivots on the diagonal.

Since the determinant of the matrix A is also needed in some applications, Brenner's subroutine SYMINV was modified to simultaneously calculate det(A). This is accomplished at each stage of the Gaussian elimination by accumulating the product of the pivots, when the chosen

pivot divides into a given row to be multiplied by a scale and subtracted from other rows.

It was found that underflows occurred in accumulating the product of the pivots, even with the automatic scaling described in Section D.3. Since ln[det(A)] is desired in maximum likelihood estimation in Chapter 7, the code was changed to accumulating the natural logarithms of the pivot elements.

D.3 Automatic Scaling

Suppose we had to solve the equations AX = B and simultaneously invert the symmetric matrix A. There could be numerical problems if, for instance, the units of the X_j were badly out of scale relative to each other (such as meters versus micro-inches, etc.). Either the units of the physical variables in a problem could be appropriately chosen, or they can be chosen in any arbitrary convenient fashion, and an automatic way of scaling the equations applied, the equations solved and the matrix inverted, and then the inverse scaling applied.

Suppose we changed units to obtain variables $X'_j = s_j X_j$. There results a new set of linear equations A'X' = B', where

$$A'_{ij} = \frac{1}{s_i s_j} A_{ij}$$

$$B'_i = \frac{1}{s_i} B_i$$

With A being a Fisher information type of matrix, it turns out that an appropriate automatic scaling choice is [3], [15]

$$s_j = \sqrt{A_{jj}}$$

With this scaling, it has been found that very large matrices can be inverted and equations solved without underflows or overflows on the computer. A scaling subroutine is placed between the SYMINV symmetric

matrix inversion and solution subroutine to automatically scale the equation $AX = B$, call SYMINV, and from the solution vector X' calculate $X_j = \frac{1}{s_j} X'_j$.

D.4 Checking Partial Derivatives by the Difference Method

In addition to the complications of, e.g., Kalman filter code, the algorithms of this thesis required the coding of many partial derivatives. A mechanical way of verifying the partial derivative code by the difference method was applied [3].

Namely, let α_{i0} and α_{i1} be two values of a parameter. For a given quantity X it was checked numerically that

$$\frac{1}{2}\left[\frac{\partial X}{\partial \alpha_i}\Big|_{\alpha_i=\alpha_{i1}} + \frac{\partial X}{\partial \alpha_i}\Big|_{\alpha_i=\alpha_{i0}}\right] = \frac{1}{\alpha_{i1}-\alpha_{i0}}\left[X\Big|_{\alpha_i=\alpha_{i1}} - X\Big|_{\alpha_i=\alpha_{i0}}\right]$$

When errors had been eliminated with this check, the estimation code was generally able to work correctly.

References

1. Aitken, A. C., *Determinants and Matrices,* Interscience Publishers Inc., New York, 1956.

2. Arnold, Ludwig, *Stochastic Differential Equations: Theory and Applications,* John Wiley & Sons, New York, 1974.

3. Ash, Michael E., "Determination of Earth Satellite Orbits," Technical Note 1972-5, M.I.T. Lincoln Laboratory, Lexington, MA, April 1972.

4. Barton, Richard J. and H. Vincent Poor, "Signal Detection in Fractional Gaussian Noise," *IEEE Transactions on Information Theory,* Vol. 34, No. 5, September 1988, pp. 943-959.

5. Bendat, Julius S. and Allan G. Piersol, *Engineering Applications of Correlation and Spectral Analysis,* John Wiley & Sons, Inc., New York, 1980.

6. Bendat, Julius S. and Allan G. Piersol, *Random Data: Analysis and Measurement Procedures,* John Wiley & Sons, Inc., New York, 1971.

7. Breiman, Leo, *Statistics: With a View Toward Applications,* Houghton Mifflin Co., Boston, 1973.

8. Brown, Robert G., *Introduction to Random Signal Analysis and Kalman Filtering,* John Wiley & Sons, Inc., New York, 1983.

9. Dieudonné, J., *Foundations of Modern Analysis,* Academic Press, New York, 1960.

10. Doob, S. L., "Weiner's Work in Probability Theory," *Bulletin of the American Mathematical Society,* Vol. 72, No. 1, Part II, January 1966, pp. 69-72.

11. Eshbach, Ovid W. and Mott Souders (ed.), *Handbook of Engineering Fundamentals,* John Wiley & Sons, New York, 1975.

12. Feller, William, *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, Inc., New York, 1950.

13. Friedman, Avner, *Stochastic Differential Equations and Applications: Volume 1*, Academic Press, Inc., New York, 1975.

14. Gelb, Arthur (ed.), *Applied Optimal Estimation*, The M.I.T. Press, Cambridge, 1974.

15. Gupta, Narendra, K. and Raman K. Mehra, "Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations," *IEEE Transactions on Automatic Control*, Vol. AC-19, No. 6, December 1974, pp. 774-783.

16. Heinemann, Peter C., Michael B. Kennedy, and John T. Prohaska, "System Identification and Model Validation Using FIMLOF (Full Information Maximum Likelihood via Optimal Filtering), R-1090, C. S. Draper Laboratory, Cambridge, MA, November 1978.

17. Hestrom, Carl W., *Probability and Stochastic Processes for Engineers*, Macmillan Publishing Company, New York, 1984.

18. Jazwinski, Andrew H., *Stochastic Processes and Filtering Theory*, Academic Press, Inc., San Diego, CA, 1970.

19. Kac, M., "Weiner and Integration in Function Spaces," *Bulletin of the American Mathematical Society*, Vol. 72, No. 1, Part II, January 1966, pp. 52-68.

20. Karatzas, Ioannis, and Steven E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer Verlag, New York, 1988.

21. Keshner, Marvin S., "1/f Noise," *Proceedings of the IEEE*, Vol. 70, No. 3, March 1982, pp. 212-218.

22. Koenigsburg, William D., "Spectral Analysis of Random Signals - Techniques and Interpretation," E-2771, C.S. Draper Laboratory, Cambridge, MA, June 1973.

23. Ljung, Lennart, "The Extended Kalman Filter as a Parameter Estimator for Linear Systems," Report LiTH-ISY-I-0154, Dept. of Electrical Engineering, Linköping University, Linköping, Sweden, 1977.

24. Lundahl, Torbjörn, William J. Ohley, Steven M. Kay and Robert Siffert, "Fractional Brownian Motion: A Maximum Likelihood Estimator and Its Application to Image Texture," *IEEE Transactions on Medical Imaging*, Vol. MI-5, No. 3, September 1986, pp. 152-161.

25. Mandlebrot, Benoit B., *The Fractal Geometry of Nature*, W. H. Freeman and Company, San Francisco, 1982.

26. Mandelbrot, Benoit B. and John W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, Vol. 10, 1969, pp. 422-437.

27. Mandlebrot, Benoit B. and James R. Wallis, "Some Long-run Properties of Geophysical Records," *Water Resources Research*, Vol. 5, No. 2, 1969, pp. 321-340.

28. Maybeck, Peter S., "Combined State and Parameter Estimation for On-Line Applications," Doctoral Dissertation, M.I.T., Cambridge, MA, February 1972; also C.S. Draper Laboratory T-557.

29. Maybeck, Peter S., *Stochastic Models, Estimation, and Control: Vol 2*, Academic Press, Inc., San Diego, CA, 1982.

30. McShane, E. J., *Stochastic Calculus and Stochastic Models*, Academic Press, Inc., New York, 1974.

31. Musicus, Bruce, R., "An Iterative Technique for Maximum Likelihood Estimation with Noisy Data," Master's Thesis, M.I.T., Cambridge, MA, February 1979.

32. Oldham, K. B., and J. Spanier, *The Fractional Calculus: Theory and Applications of Differentiation and Integration of Arbitrary Order*, Academic Press, New York, 1974.

33. Papoulis, Athanasios, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Co., New York, 1984.

34. Papoulis, Athanasios, *Signal Analysis*, McGraw-Hill Book Company, New York, 1977.

35. Parzen, Emanuel, *Stochastic Processes*, Holden-Day, Inc., San Francisco, 1962.

36. Radeka, Veljko, "1/|f| Noise in Physical Measurements," *IEEE Transactions on Nuclear Science,* Vol. NS-16, pp. 17-35, October 1969, pp. 178-196.

37. Sandel, Nils R. Jr. and Khaled I. Yared, "Maximum Likelihood Identification of State Space Models for Linear Dynamic Systems," M.I.T. Electronic Systems Laboratory Report ESL-R-814, Cambridge, MA, April 1978.

38. Schweppe, Fred C., "Algorithms for Estimating a Re-Entry Body's Position, Velocity, and Ballistic Coefficient in Real Time or for Post Flight Analysis," Report 1964-4, M.I.T. Lincoln Laboratory, Lexington, MA, January 1965.

39. Schweppe, Fred C., "Evaluation of Likelihood Functions for Gaussian Signals," *IEEE Trans. on Information Theory,* Vol. IT-11, No. 1, January 1965, pp. 61-70.

40. Strang, Gilbert, *Introduction to Applied Mathematics,* Wellesley-Cambridge Press, Wellesley, MA, 1986.

41. Taylor, Robert M., "Dynamic System Identification Based on Propagation of Probability Densities and Their Partial Derivatives," CSDL-P-1929, C. S. Draper Laboratory, Cambridge, MA, July 1984.

42. Todoeschuck, John P. and Oliver G. Jensen, "Joseph Geology and Seismic Deconvolution," *Geophysics,* Vol. 53, No. 11, November 1988, pp. 1410-1414.

43. Voss, Richard F. "1/f (Flicker) Noise: A Brief Review," Proc. 33rd Annual Symp. on Freq. Contr., Atlantic City, NJ, 1979, pp. 40-46.

44. West, Bruce J. and Michael Shlesinger, "The Noise in Natural Phenomena," *American Scientist,* Vol. 78, Jan.-Feb. 1990, pp. 40-45.

45. Wilks, Samuel, S., *Mathematical Statistics,* John Wiley & Sons, Inc., New York, 1962.